Studies of eucalypt genetic and genomic architecture

Jakob Burton Butler

B. Sc (Hons)

Submitted in fulfilment of the requirements for the Doctorate of Philosophy

School of Natural Sciences, University of Tasmania

May, 2018

Declarations

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

This thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act 1968.

The publishers of the papers comprising Chapters 2 to 4 hold the copyright for that content and access to the material should be sought from the respective journals. The remaining non published content of the thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act 1968.

Signature: Date: 14/05/2018

Jakob B. Butler, Candidate

We the undersigned agree with the above declarations, and approve the submission of the final corrected version of this thesis:

René E. Vaillancourt

Supervisor
ARC Centre for Forest Value and School of Natural Sciences
University of Tasmania

Signed:

Date: 10/05/2018

Brad M. Potts

Supervisor
ARC Centre for Forest Value and School of Natural Sciences
University of Tasmania

Signed:

Date: 8/5/2018

Jules S. Freeman

Supervisor
ARC Centre for Forest Value and School of Natural Sciences
University of Tasmania &
Scion, Rotorua. New Zealand

Signed

Date: 14/05/2018

Mark A. Hunt

Head of School
School of Natural Sciences

University of Tasmania

Date:

Statement of co-authorship/contribution

The following people and institutions contributed to the publication of work undertaken as part of this thesis:

Jakob B. Butler, School of Natural Sciences, University of Tasmania, Hobart, Australia = **Candidate**

René E. Vaillancourt, School of Natural Sciences and ARC Centre for Forest Value, University of Tasmania, Hobart, Australia = **REV**

Brad M. Potts, School of Natural Sciences and ARC Centre for Forest Value, University of Tasmania, Hobart, Australia = **BMP**

Jules S. Freeman, ARC Centre for Forest Value, University of Tasmania, Hobart, Australia and Scion, Rotorua, New Zealand= **JSF**

David J. Lee, Forest Industries Research Centre, University of the Sunshine Coast, Maroochydore DC, Australia = **DJL**

Graham J. King, Southern Cross Plant Science, Southern Cross University, Lismore, Australia = **GJK**

Abdul Baten, Southern Cross Plant Science, Southern Cross University, Lismore, Australia = **AB**

Mervyn Shepherd, Southern Cross Plant Science, Southern Cross University, Lismore, Australia = **MS**

Morag Glen, Tasmanian Institute of Agriculture, University of Tasmania, Hobart, Australia = **MG**

Geoff S. Pegg, Department of Agriculture, Fisheries and Forestry, Brisbane, Australia = **GSP**

Dario Grattapaglia, EMBRAPA Genetic Resources and Biotechnology, Brasilia, Brazil = **DG**

Orzenil B. Silva-Junior, EMBRAPA Genetic Resources and Biotechnology, Brasilia, Brazil = **OBS**

Blake A. Simmons, DOE Joint Bioenergy Institute, Lawrence Berkeley National Laboratory, Berkeley, USA = **BAS**

Adam L. Healey, DOE Joint Bioenergy Institute, Lawrence Berkeley National Laboratory, Berkeley, USA = **ALH**

Jeremy Schmutz, Hudson-Alpha Institute for Biotechnology, Huntsville and DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, USA = **JS**

Kerrie W. Barry, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, USA = **KWB**

Robert J. Henry, University of Queensland and Queensland Alliance for Agriculture and Food Innovation, Brisbane, Australia = **RJH**

Below are the details of co-authors for each paper/thesis chapter and their contributions:

Chapter 2 is published as:

Butler JB, Vaillancourt RE, Potts BM, Lee DJ, King GJ, Baten A, Shepherd M, Freeman JS (2017) Comparative genomics of *Eucalyptus* and *Corymbia* reveals low rates of genome structural rearrangement. BMC Genomics 18:397

Candidate performed marker analysis, linkage map construction and was the primary author. All authors contributed to the experimental design of the study and critically reviewed the manuscript; crossing of pedigrees and subsequent monitoring of growth was performed by DJL; collection of samples and DNA extraction was performed by Candidate and MS; JSF provided guidance on linkage map construction.

Chapter 3 is published as:

Butler JB, Freeman JS, Vaillancourt RE, Potts BM, Glen M, Lee DJ, Pegg GS (2016) Evidence for different QTL underlying the immune and hypersensitive responses of *Eucalyptus globulus* to the rust pathogen *Puccinia psidii*. Tree Genetics & Genomes 12:1-13

Candidate performed QTL analysis and was the primary author. All authors critically reviewed the manuscript; Candidate and REV, BMP, JSF, DJL, GSP contributed to the experimental design; crossing of pedigrees and subsequent monitoring of growth was performed by DJL; inoculation and disease assessment was performed by GSP; identification of *E. globulus* pathogens was performed by MG; JSF provided guidance on QTL analysis.

Chapter 4 is published as:

Butler JB, Freeman JS, Potts BM, Vaillancourt RE, Grattapaglia D, Silva-Junior OB, Simmons BA, Healey AL, Schmutz J, Barry KW, Lee DJ, Henry RJ, King GJ, Baten A, Shepherd M (2018) Annotation of the *Corymbia* terpene synthase gene family shows broad conservation but dynamic evolution of physical clusters relative to *Eucalyptus*. Heredity

Candidate performed manual gene annotation and was the primary author. All authors contributed to the experimental design of the study and critically reviewed the manuscript; the *Corymbia* genome assembly was performed by AB, OBS, ALH with assistance from GJK, MS, DG, BAS, JS, HWB; MS provided guidance on gene annotation.

Chapter 5 is in preparation for publication as:

Butler JB, Vaillancourt RE, Potts BM, Pegg GS, Lee DJ, Freeman JS (in prep) Independent QTL for resistance to the native pathogen *Quambalaria pitereka* and the exotic pathogen *Austropuccinia psidii* in *Corymbia.*

Candidate performed QTL analysis and was the primary author. All authors contributed to the experimental design and critically reviewed the manuscript; crossing of pedigrees and subsequent monitoring of growth was performed by DJL; inoculation and disease assessment was performed by GSP.

We the undersigned agree with the above stated "proportion of work undertaken" for each of the above published peer-reviewed (or in preparation) manuscripts contributing to this thesis:

René E. Vaillancourt

Supervisor

ARC Centre for Forest Value and School of Natural Sciences

University of Tasmania

Signed:

Date: 8/5/2018

Brad M. Potts

Supervisor

ARC Centre for Forest Value and School of Natural Sciences

University of Tasmania

Signed:

Date: 8/5/2018

Jules S. Freeman

Supervisor

ARC Centre for Forest Value and School of Natural Sciences

University of Tasmania &

Scion, Rotorua, New Zealand

Signed

Date: 8/5/2018

Mark A. Hunt

Head of School

School of Natural Sciences

University of Tasmania

Signed:

Date:

Jakob B. Butler

Candidate

School of Natural Sciences

University of Tasmania

Signed:

Date: 14/05/2018

Abstract

The utility of genomics for theoretical and practical outcomes has been greatly enhanced by the advent of high throughput molecular technologies. This thesis reports the application of linkage mapping, quantitative trait loci analysis, gene family annotation and comparative genomics to study various aspects of the genetic and genomic architecture of eucalypts, a group of Myrtaceous flora that are of significant economic and ecological value worldwide.

Linkage mapping was employed to compare the genomic architecture of *Corymbia* and *Eucalyptus.* Three independent high density linkage maps for two *Corymbia* species (*Corymbia citriodora* subsp. *variegata* and *C. torelliana*) were constructed from two hybrid pedigrees of *C. torelliana* individuals crossed with a common *C. citriodora* subsp. *variegata* parent. Subsequent analysis provided evidence for large (from 1 - 13 MB) intra-chromosomal rearrangements between *Corymbia* and *Eucalyptus* on seven of their 11 chromosomes. Most rearrangements were supported through comparisons of the three *Corymbia* maps to the *E. grandis* reference genome, and to other *Eucalyptus* linkage maps. These are the first large-scale chromosomal rearrangements discovered between eucalypts. However, in the context of a divergence approximately 52 million years ago, the genomic structure of the two genera was remarkably conserved; adding to growing evidence for conservation of genome structure within lineages of woody angiosperms. These maps informed the collaborative assembly of the *Corymbia* reference genome (not itself part of this thesis), which formed the basis of further research reported below.

The genetic basis of variation in resistance to myrtle rust, caused by *Austropuccinia psidii* (formerly *Puccinia psidii*), was examined using an *E. globulus* linkage map. Quantitative trait loci (QTL) analysis was undertaken using 218 genotypes of an outcross *E. globulus* F_2 mapping family, phenotyped by controlled inoculation of their open pollinated progeny (11 per genotype on average) with *A. psidii*. To examine possible independent control of different aspects of plant resistance, QTL analyses were conducted by classifying individuals as symptomless *versus* those exhibiting symptoms, and those exhibiting a hypersensitive reaction *versus* more severe symptoms. Four QTL were identified; two influencing the symptomless response, and two influencing the hypersensitive response. The potential resistance mechanisms underlying these different QTL are discussed. Together with past findings, this study suggests that *A. psidii* resistance in eucalypts is quantitative in nature and influenced by the complex interaction of multiple loci of variable effect.

To examine differences in genetic architecture underlying disease resistance in *Corymbia* and *Eucalyptus* and explore differences in resistance to exotic and co-evolved pathogens,

the *Corymbia* linkage maps were used to perform the first QTL study for disease resistance in this genus. Resistance was examined to the pandemic strain of *A. psidii* and two strains of the native pathogen *Quambalaria pitereka* (QSB1 and QSB2). These analyses were undertaken using 360 genotypes from the two *C. citriodora* subsp. *variegata* x *C. torelliana* F_1 hybrid mapping crosses, phenotyped in separate controlled inoculations. Twenty QTL were identified; six for rust, nine for QSB1 and five for QSB2. Positioning these QTL on the *Corymbia* reference genome revealed only one case of QTL co-location (peak location within ± 2 MB) between rust and QSB, while no QTL for either of the QSB strains were co-located. Resistance to *A. psidii* and *Q. pitereka* in *Corymbia* appears to be controlled by multiple independent loci, with a larger percentage of variation explained by both the mean effect of individual QTL and the total combined effect within pedigree in the response to QSB compared to rust. Notable co-locations with *E. globulus* resistance QTL for rust and other pathogens were detected, and the implications of this conservation explored.

Variation in genome architecture between *Corymbia* and *Eucalyptus* was further examined through comparison of the number, phylogenetic relatedness and physical distribution of the *terpene synthase* (*TPS)* gene family. Terpenes are important foliar chemicals for both primary and secondary metabolism, and *Eucalyptus* is notable for its expansive *TPS* gene family. This gene family was manually annotated in the *Corymbia* reference genome, revealing a similar overall number and subfamily representation of *TPS* genes relative to *Eucalyptus*, suggesting these features are characteristic of eucalypts. Physical arrangement of *TPS* genes involved in the synthesis of secondary metabolites differed significantly between *Eucalyptus* and *Corymbia* with translocation, expansion/contraction and loss of *TPS* gene clusters. In contrast, those involved in primary metabolism were often highly conserved, likely reflecting different selective constraints. The mechanisms underlying the fine-scale variability in *TPS* genes despite the broad conservation observed between the eucalypts are explored.

Combined, these results reveal a common theme of broad conservation in genomic and genetic architecture between different eucalypt genera, with greater variation in fine-scale features such as chromosomal structure, genetic architecture underlying variation in disease resistance and gene family arrangement. This work adds to a fundamental understanding of what differentiates *Eucalyptus* and *Corymbia*, and serves to highlight the potential differences one can expect between tree genera.

Acknowledgements

I would like to thank my supervisory team of René Vaillancourt, Jules Freeman and Brad Potts for their stimulating discussions, humour, scepticism and support of all aspects of this PhD, without which this thesis would have been all the poorer.

I would also like to thank our collaborators in the *Corymbia* genome consortium for the opportunity to be part of such an important undertaking. In particular, I'd like to thank Graham King, Abdul Baten and Mervyn Shepherd for hosting me twice at Southern Cross University over the course of this period, and for all the extensive consultation on working with the genome assembly.

I wish to also acknowledge the financial support provided by the Tasmanian Graduate Research Scholarship, and also the School of Biological Science for their contributions to conference travel costs.

To my peers, colleagues and office-mates in the School of Biological/Natural Sciences, the culture and comradery that you all bring makes coming to university a pleasure.

To my parents Philip and Tonia, thanks for bankrolling my academic career to this point. Here's hoping that you see a return on this investment at some point.

Shout out to my son Finnegan, whose birth has made the last few months of my PhD all the more exciting.

Final thanks must go to my partner Catherine, whose relentless cheerfulness, enthusiastic support and bemused attempts at understanding my work have helped to make the last few years the most rewarding of my life.

List of publications and presentations arising from this thesis

Refereed journal articles:

Butler JB, Freeman JS, Vaillancourt RE, Potts BM, Glen M, Lee DJ, Pegg GS (2016) Evidence for different QTL underlying the immune and hypersensitive responses of *Eucalyptus globulus* to the rust pathogen *Puccinia psidii*. Tree Genetics & Genomes 12:1-13

Butler JB, Vaillancourt RE, Potts BM, Lee DJ, King GJ, Baten A, Shepherd M, Freeman JS (2017) Comparative genomics of *Eucalyptus* and *Corymbia* reveals low rates of genome structural rearrangement. BMC Genomics 18:397

Butler JB, Freeman JS, Potts BM, Vaillancourt RE, Grattapaglia D, Silva-Junior OB, Simmons BA, Healey AL, Schmutz J, Barry KW, Lee DJ, Henry RJ, King GJ, Baten A, Shepherd M (2018) Annotation of the *Corymbia* terpene synthase gene family shows broad conservation but dynamic evolution of physical clusters relative to *Eucalyptus*. Heredity

Calvert J, Baten A, Butler J, Barkla B, Shepherd M (2018) Terpene synthase genes in *Melaleuca alternifolia*: comparative analysis of lineage-specific subfamily variation within Myrtaceae. Plant Systematics and Evolution 304:111-121

Conference presentations

Shepherd M, Baten A, Junior OBdS, Lee DJ, Butler JB, Freeman J, Vaillancourt R, Potts B, Grattapaglia D, King G (2015) Towards a *Corymbia* reference genome: comparative efficiencies of Illumina, PacBio and hybrid *de novo* assemblies of a complex heterozygous genome. In: Vettori C, Vendramin GG, Paffetti D, Travaglini D (eds) Proceedings of the IUFRO Tree Biotechnology 2015 Conference: "Forests: the importance to the planet and society", Florence, Italy.

Shepherd M, Barry KW, Baten A, Butler JB, Freeman JS, Furtado A, Grattapaglia D, Healey A, Henry RJ, King GJ, Lee D, Potts BM, Schmutz J, da Silva Junior OB, Simmons B, Vaillancourt R (2016) Spotting the difference: comparing the genome of *Corymbia* with its larger cousin *Eucalyptus grandis*. In: Plant & Animal Genome Conference XXIV, San Diego, CA.

Butler JB, Freeman JS, Potts BM, Vaillancourt RE, Grattapaglia D, Junior OBdS, Simmons B, Schmutz J, Barry KW, Lee DJ, Healey A, Furtado A, Henry RJ, Baten A, King GJ, Shepherd M (2016) Comparing the terpene synthase gene family in two assembled eucalypt genomes. In: Phylomania 2016, Hobart, Australia.

Healey A, Shepherd M, Baten A, King GJ, Lee DJ, Furtado A, Vaillancourt RE, Butler JB, Freeman JS, Potts BM, Grattapaglia D, Junior OBdS, Barry KW, Schmutz J, Simmons B, Henry RJ (2017) Sequencing the branches of the eucalypt tree: comparison between *Eucalyptus* and *Corymbia* genomes. In: Plant & Animal Genome Conference XXV, San Diego, United States of America.

Posters

Butler J, Freeman J, Potts B, Lee D, King G, Baten A, Shepherd M, Vaillancourt R (2016) Structural differences between the genomes of the major eucalypt lineages. In: Forest Genetics for Productivity Conference, Rotorua, New Zealand.

Butler J, Freeman J, Vaillancourt R, Potts B, Glen M, Lee D, Pegg G (2016) Symptomless and hypersensitive resistance of *Eucalyptus globulus* to the rust pathogen *Puccinia psidii*: evidence for different QTL underlying each response. In: Forest Genetics for Productivity Conference, Rotorua, New Zealand.

Butler J, Freeman J, Potts B, Vaillancourt R, Grattapaglia D, da Silva Junior O, Simmons B, Schmutz J, Barry K, Lee D, Healey A, Furtado A, Henry R, Baten A, King G, Shepherd M (2017) Comparative analysis of the terpene synthase gene family between *Eucalyptus* and *Corymbia*. In: Plant & Animal Genome Conference XXV, San Diego, United States of America.

Table of contents

Declarations ..... **i**

Statement of co-authorship/contribution ..... **ii**

Abstract ..... **vi**

Acknowledgements ..... **viii**

List of publications and presentations arising from this thesis ..... **ix**

Table of contents ..... **xi**

Chapter 1 - General introduction **1**

Genetic versus genomic architecture 1

The eucalypts: genera of international importance 3

Linkage mapping as a genomic tool 4

The advent of the genome assembly 6

Applications of genome assemblies 8

Thesis outline 10

Chapter 2 - Comparative genomics of *Eucalyptus* and *Corymbia* reveals low rates of genome structural rearrangement **11**

Introduction 11

Material and methods 14

Genetic material 14

DNA extraction protocol 14

DArTseq genotyping 15

Linkage map construction 15

Parental bin map construction 16

Comparison of parental bin maps 17

Mapping of CCV using segregation data from the two different populations, and comprehensive map construction 18

Comparison with *Eucalyptus grandis* 19

Results 19

DArTseq genotyping 19

Comparison of parental bin maps 19

Comprehensive maps 20

Collinearity of *Corymbia citriodora* subsp. *variegata* with *Eucalyptus grandis* 22

Discussion 25

Supplementary material 30

Chapter 3 - Evidence for different QTL underlying the immune and hypersensitive responses of *Eucalyptus globulus* to the rust pathogen *Puccinia psidii* 33

Introduction 33

Material and methods 35

Experimental design 35

Inoculum, inoculation and rust assessment 36

QTL analysis for rust 37

QTL analysis for other fungal diseases 38

Positioning of candidate genes and other QTL 39

Results 39

Analysis of variation in rust resistance 39

QTL for disease resistance 41

Epistatic QTL 45

Candidate genes/co-location 45

Discussion 46

Supplementary material 52

Chapter 4 - A comparison of QTL for resistance to the native pathogen *Quambalaria pitereka* and the exotic pathogen *Austropuccinia psidii* in *Corymbia* 66

Introduction 66

Material and methods 70

Genetic material, inoculation and symptoms assessment 70

QTL analysis 71

Positioning of QTL for other pathogens and candidate genes 72

Results 73

Variation in rust and QSB resistance 73

QTL discovery and effect size 75

Co-location of QTL from different crosses using the *Corymbia* reference genome 78

Candidate gene discovery 79

Discussion 79

Differing QTL resistance effect size between native and exotic pathogens 80

Notable co-locations between QTL 81

Positional candidate genes associated with resistance 82

Strain specificity of resistance responses 83

Conclusions 84

Supplementary material 85

Chapter 5 - Annotation of the *Corymbia* terpene synthase gene family shows broad conservation but dynamic evolution of physical clusters relative to *Eucalyptus*..... 95

Introduction 95

Material and methods 97

Terpene synthase gene discovery..... 97

Phylogenetic analysis and annotation 98

RNA-Seq expression analysis 99

Comparative analysis of the *TPS* gene family between species 99

Results..... 100

Discovery of *TPS* loci 100

Phylogenetic analysis..... 101

Proportional representation and genome organisation of *TPS* genes 106

TPS gene expression 111

Discussion 114

Broad conservation in the eucalypt *TPS* family 114

Variation in the *TPS* genes specific to each eucalypt lineage 115

Variation in the *TPS* subfamilies involved in secondary metabolite synthesis..... 117

Conservation in the *TPS* subfamilies involved in primary metabolite synthesis 117

Contributions of stochastic and selective pressures to the variation in the *TPS* gene family 118

Pseudogenes and expression of *TPS* loci 119

Physical structure of *TPS* gene clusters 120

Conclusions..... 121

Supplementary material 122

Chapter 6 - General discussion 134

Broad genomic conservation between the eucalypt genera, consistent with other forest trees..... 134

Fine-scale genomic and genetic differences between eucalypt genera..... 135

Application of these genomic resources..... 137

Conclusions..... 137

References 139

Chapter 1 - General introduction

The study of genetic and genomic architecture for theoretical and practical outcomes has been greatly enhanced by the advent of high throughput molecular technologies, along with improvements in computing power and analytical techniques. Historically, the study of entire genomes was largely limited to what was physically visible via karyotype analysis, however the vast increases in available data due to the development of high throughput DNA-based technologies has now made possible a plethora of new areas of research. High density linkage mapping (Groenen et al. 2009; Berlin et al. 2010; Bartholomé et al. 2015), genome sequencing (Nystedt et al. 2013; Myburg et al. 2014) and genomic selection (Hayes et al. 2009; Jannink et al. 2010) have all become relatively commonplace, and as technology improves and becomes more cost effective, the applications of genomics will continue to progress. This chapter will explore the various genomic and genetic techniques used for the study of genomes, with special attention directed to their use in eucalypts.

Genetic versus genomic architecture

The study of genetic architecture, in essence, seeks to quantify the relationship between genotype and phenotype (Hansen 2006). Commonly studied aspects of genetic architecture include the heritability of traits (especially narrow-sense heritability, the contribution of additive genetic effects on phenotype), the effects of dominance, epistasis or pleiotropy, and the number of genes and alleles affecting a phenotype (Hansen 2006), permitting investigation into broader concepts such as species and population differentiation (Fenster and Galloway 2000; Zeng et al. 2000), processes of adaptation and selection (Laurie et al. 2004; Yeaman and Whitlock 2011), and the genetic control of various traits (Mackay 2001). Since the earliest studies of genetic architecture (Fisher 1930), advancements in this field have led to the formation of a branch of genetics called quantitative genetics, which has been important to our understanding of the genetic basis of phenotypic variation.

Technological advances in the last few decades now allow studies of genetic architecture to encompass the entirety of the genetic information contained in an individual, expanding these studies to a genomic level. Studies of genomic architecture, defined as the totality of the non-random arrangement of functional elements in a genome (Koonin 2009), are distinguished from genetic architectural studies by the scale of investigation, seeking to examine the position and arrangement of all features within a genome. These studies are driven by the understanding that the phenotype of an individual is influenced not just by the sum of genetic effects, but also by the arrangement and context of these parts (Lynch

2007). These features can extend from the (relatively) small-scale arrangement of consecutive genes, to chromosome and genome-wide structuring of sequence, motifs and functional elements.

A notable and well-studied example of genomic architecture is the occurrence of prokaryotic genes organized into co-transcribed groups called operons (Wilson et al. 2007). The original description of a cluster of genes regulated by a local repressor (Jacob and Monod 1961) stimulated a great deal of study into the occurrence of this gene arrangement in bacteria. While gene order across different bacterial species is rarely conserved (Koonin et al. 1996; Dandekar et al. 1998), genes coding for proteins that physically interact with each other are often arranged in coherent blocks (operons) and are common to almost all bacterial species (Koonin 2009). This arrangement is advantageous due to the efficiency of expressing related products through a single regulatory mechanism, and is one of the defining features of genome architecture in prokaryotic organisms (Salgado et al. 2000). While operons are (with rare exceptions) not found in eukaryotic genomes (Lee and Sonnhammer 2003), there is still widespread gene organization into both functionally similar and/or co-expressed groups, most of which are genera- or species-specific (Chu et al. 2011; Papantonis and Cook 2013). This organisation is believed to be influenced by the interplay of various mechanisms. For instance, tandem duplication is a common mechanism influencing many gene families, resulting in large arrays of clustered genes with similar function (Cannon et al. 2004). Homologous recombination can also result in genes from related metabolic pathways becoming locally clustered (Leister 2004; Hanada et al. 2008). However, the most common fate of the products of tandem duplication and homologous recombination is gene elimination through purifying selection. The strength of this purifying selection is proportional to the effective size of the population, so altered genome structure or increased complexity can often be traced to genetic bottlenecks (Koonin 2009). The study of genomic architecture is therefore essential to develop an understanding of the evolutionary history underlying these arrangements and any advantages they provide.

With the availability of new data generated from analyses of genomic architecture, many novel investigations are now possible. For instance, there is now much greater scope for comparative genomic analyses between related organisms, such as the study of variation in gross genomic architecture via markers of known genomic location or sequence (Lyons et al. 2008; Hudson et al. 2012b; Li et al. 2015). This can reveal the presence of distinct inter- or intra-chromosomal features such as inversions or translocations between a pair of species, hinting at the shared and unique events in their evolutionary history (Hanley et al. 2006; Luo et al. 2015). Genomic architecture is also commonly studied within species, such

as examining the arrangement of genes associated with disease resistance (Gururani et al. 2012) and performing genome wide association studies (Korte and Farlow 2013). These genomic investigations further our understanding of the underlying complexity in the genome of all organisms and how this relates to phenotypic variation.

The eucalypts: genera of international importance

Organisms of ecological and economic importance are often prioritised for genomic studies. In Australia, one of the most important groups of flora are the Myrtaceae, a predominantly southern hemisphere woody plant family containing over 5,650 species in 130 to 150 different genera (Grattapaglia et al. 2012). This family is spread over the entire Australian continent and is a major component of most native plant communities (Myerscough 1998). As well as their ecological importance, many genera of Myrtaceae are also economically valuable globally, underpinning entire industries including essential oils, wood pulp and timber (Zobel 1993; Doughty 2000; Homer et al. 2000; Batish et al. 2008; Hamilton et al. 2008).

Of particular importance are the eucalypts, a group of Myrtaceae including the genera *Eucalyptus, Corymbia* and *Angophora*, with over 900 different species recognised (Brooker 2000; Slee et al. 2006). While mainly native to Australia, their fast growth rate, wood quality and tolerance for varied climates has led to *Eucalyptus* and *Corymbia* being grown around the world for diverse purposes (Doughty 2000; Poke et al. 2005). *Eucalyptus*, for instance, are the predominant hardwood plantation species in Australia and overseas, due to their importance to the pulp, charcoal, fuel and timber industries (Doughty 2000; Rockwood et al. 2008; Shepherd et al. 2011). Likewise, *Corymbia* is increasingly cultivated for timber and essential oil production in Australia, India, Brazil, Fiji and South Africa (Asante et al. 2001; Vernin et al. 2004). Specific species are often chosen for their performance in different climates (Booth and Pryor 1991); for instance *Eucalyptus globulus* is one of the primary pulpwood plantation species in temperate regions (Gavran 2014), while in the tropical climate of Brazil *Eucalyptus grandis* and its hybrids make up the majority of plantations (Almeida et al. 2004).

In species of economic significance, research is often targeted toward understanding the genetic basis of variation in economically important traits to assist in the improvement of genotypes in breeding programs (Lee 2007; Hamilton et al. 2008). Before the advent of genomics, quantitative genetic analysis was the main approach employed in eucalypts to study the genetic architecture of traits such as product yield (Hamilton et al. 2008) and tolerance to disease and environmental stresses (Balmelli et al. 2014; Pegg et al. 2014a). As

technology improved and became more cost effective, genomic resources have been developed which contribute to an enhanced fundamental understanding of the genetic architecture of phenotypic trait variation in eucalypts and are likely to be used for genomic selection in the future (Grattapaglia and Kirst 2008; Grattapaglia and Resende 2011; Myburg et al. 2014).

Linkage mapping as a genomic tool

The basis of linkage mapping is genetic recombination resulting from crossing over between homologous chromosomes during meiosis (Sturtevant 1913). Estimating how often recombination occurs between a pair of markers in a cross allows the calculation of a genetic 'distance' between them, defined in terms of centimorgans (cM, where 1 cM = 1% recombination). This allows the creation of a genetic linkage map in which markers are positioned by their cM distance relative to adjacent markers (Sturtevant 1913). While linkage maps were originally created with phenotypic markers (Taylor and Trotter 1967) the transferability (ability to be genotyped in multiple species or individuals) of these markers was limited, and the comparison of linkage maps between species was not possible until the advent of the specific molecular markers explored below (Semagn et al. 2006). A molecular marker is representative of differences between organisms at the DNA level. The development of restriction fragment length polymorphisms (RFLPs) (Sambrook et al. 1975) led to the first published human linkage map using molecular markers (Botstein et al. 1980). With the advent of new marker technologies such as simple sequence repeats (SSRs) (Tautz and Renz 1984; Litt and Luty 1989) and single nucleotide polymorphisms (SNPs) (Williams et al. 1990) maps were created with greater marker density and higher transferability between species. Indeed, one can trace the development of linkage maps in eucalypts based on technologies as they became available; from RAPD markers (Grattapaglia and Sederoff 1994), RFLPs (Byrne et al. 1997b) and microsatellites (Brondani et al. 2006) to microarray based technologies (Kullan et al. 2011; Hudson et al. 2012a) and SNPs (Freeman 2014; Bartholomé et al. 2015), in each case improving map density and transferability between pedigrees, thus increasing the scope of their application.

As a genomic resource, linkage maps have broad utility. One of their main applications has been to elucidate the genetic architecture underlying variation in traits, using quantitative trait loci (QTL) analysis (Junghans et al. 2003a; Freeman et al. 2013). QTL analysis provides an estimation of the number, location and magnitude of effect of loci influencing variation in quantitative traits within a bi-parental mapping cross (Sewell and Neale 2000). QTL are detected by scoring a trait in a mapping cross and searching for statistically significant associations between the segregation of mapped markers and that particular trait. The

discovery of QTL can lead to marker-assisted selection in some cases, but more commonly to short-listing among potential candidate genes, and occasionally to map based cloning (Pflieger et al. 2001; Remington et al. 2001; Collard et al. 2005). They are also invaluable for gaining an understanding of the genetic architecture underlying variation in quantitative traits, suggesting whether a trait is influenced by few major effect loci, or multiple loci of smaller effect (Jansen and Stam 1994).

The precision of QTL location is dependent on several factors, including the sample size, contrasting strength of the phenotypic effect of the segregating QTL alleles and (to a smaller extent) the density of the linkage map (Lynch and Walsh 1998). As a linkage map only tracks variation in the two parents of a single cross, it is likely that the segregation of low frequency QTL alleles in a population will not be captured and therefore will go undetected (Grattapaglia and Kirst 2008), thus QTL studies are best done using multiple families (Freeman et al. 2013). QTL studies in forest trees were initially hampered by factors including long generation time and significant genetic load due to inbreeding, which limited the availability of controlled pollinated crosses. The development of the pseudo-testcross technique (Grattapaglia and Sederoff 1994) eased this constraint by allowing the use of dominant markers and inbred line approaches in two-generation outbred pedigrees. Further, QTL can be restricted to the pedigree, environment and ontogenetic stage in which they are detected (Sewell and Neale 2000), and the lack of markers that were transferable across pedigrees was a limitation for QTL validation outside the crosses used for detection. However, these limitations are beginning to be overcome as more control pollinated pedigrees, including those specifically designed to segregate for target traits, are becoming available (Freeman et al. 2008b; Hudson et al. 2014). Further, the increasing application of high throughput transferable markers (including sequence anchored markers) to the construction of linkage maps in eucalypts now allows validation of QTL locations across pedigrees, environments (Freeman et al. 2013) and ontogenic stages (Ammitzboll et al. 2018). QTL positions can also find support through co-location with SNPs from association studies (Thavamanikumar et al. 2014) and upregulated genes from expression analyses (Hudson et al. 2014). The combination of all these factors have led to the discovery of a multitude of QTL in eucalypts influencing traits such as heterochrony and flowering time (Hudson et al. 2014), growth and wood properties (Byrne et al. 1997a; Bundock et al. 2008; Freeman et al. 2009; Freeman et al. 2013), vegetative propagation (Grattapaglia et al. 1995; Marques et al. 2002), stress tolerance (Byrne et al. 1997b), fungal disease resistance (Junghans et al. 2003a; Freeman et al. 2008b), physiological disorder susceptibility (Ammitzboll et al. 2018) and foliar chemical composition (Shepherd et al. 1999; Henery et al. 2007; Freeman et al. 2008a; O'Reilly-Wapstra et al. 2011; Gosney et al. 2016).

Maps based on transferable molecular markers (such as SSRs, DArT and SNPs that are sequence based) have also enhanced the scope of comparative genomic analyses. Many of these markers are not species specific and are able to be genotyped in related species (Tautz and Renz 1984), allowing linkage maps with common markers in different organisms to be compared (Ahn and Tanksley 1993; Saghai Maroof et al. 1996; Kole et al. 2002). This can reveal differences in genomic architecture such as chromosomal inversions and translocations, and is commonly employed to elucidate differences between closely related species (Kukekova et al. 2007; Hudson et al. 2012b). The discovery of such features relies on both the density and accuracy of the linkage map and the scale of any differences between the compared species, as a more sparsely populated map may be unable to detect small chromosomal rearrangements (Chittenden et al. 1994). To ensure any features are representative of the species at large and not an individual polymorphism, validation of results through replication is also desirable (Mace et al. 2009). Sequence based markers have the added advantage of anchoring to available genome assemblies, removing the need for markers to be genotyped in both organisms and broadening the scale of comparisons able to be drawn (Jaillon et al. 2007). Linkage maps created with sequence based markers can also be used to inform the construction of genome assemblies (International Barley Genome Sequencing Consortium 2012; Kawakami et al. 2014; Bartholomé et al. 2015), and are an effective tool to employ when a completely *de novo* assembly isn't possible (explored below).

The advent of the genome assembly

Alongside the developments in molecular marker technology, techniques for determining the nucleic acid sequence of DNA were also advancing. The first widely used method involved sequencing via enzymatic polymerization using fluorescently labelled oligonucleotides, termed Sanger sequencing (Sanger et al. 1977). The advent of high-throughput sequencing a decade ago involving many sequencing reactions in parallel (Reuter et al. 2015) has caused an exponential growth in the amount of genomic information able to be investigated (Howe et al. 2008). Currently there are several different sequencing technologies available, and as each has advantages and disadvantages in read length, accuracy, cost and time, the technique chosen is often based on the intended applications (Reuter et al. 2015).

With the advent of DNA sequencing, the ambition to sequence and assemble the whole genome of an organism began to develop (Olson 1993). The most high profile project of this nature was the Human Genome project, spanning 10 years, thousands of scientists and

billions of dollars (Venter et al. 2001). This was a massive undertaking as it generally employed Sanger sequencing, with the inherent limitations in speed, cost and read length. One of the defining characteristics of Sanger sequencing (and indeed, of all sequencing technologies) is the upper limit to the length of a DNA strand which can be sequenced in a single reaction (Church and Gilbert 1984), as no current technology is able to sequence a chromosome-length DNA strand from start to finish. Instead, one must attempt to capture the entire genome represented in fragments of 50 - 15,000 bp, depending on the technique used (Reuter et al. 2015). With sufficient depth to overcome the inherent error rate in these technologies, one is able to construct contiguous stretches of DNA (contigs) from these smaller reads, often to the scale of significant portions of a chromosome (Nagarajan and Pop 2013).

The assembly of sequence into a coherent genome assembly is a demanding undertaking, for many reasons. The construction of contigs has high computational costs, often requiring weeks of calculation on high performance computing clusters to pour through the billions of base pairs of sequence data (Nagarajan and Pop 2013). There are also practical challenges to overcome caused by the characteristics of both the DNA and sequencing reactions. For instance, areas of high GC content are often under-represented during sequencing, due to the higher stability of these fragments causing them to be less preferentially amplified during PCR (Benjamini and Speed 2012). Highly repetitive DNA is often difficult to assemble, as no single read may be able to capture the full extent of the repeated area, therefore leading to similar reads (from different loci) being assigned to the same place during assembly (Treangen and Salzberg 2012). Highly heterozygous DNA can also pose a challenge in creating a single contiguous pathway through the genome. This is particularly a problem in outcrossed organisms with high genetic diversity and a short history of domestication, such as forest trees (Kajitani et al. 2014). Strategies can be employed to minimise the impact of these limitations (Reuter et al. 2015), such as specific experimental design (inbreeding to reduce heterozygosity) and various sequencing techniques (such as long reads extending over repetitive regions).

If insufficient depth and coverage is achieved to assemble the sequence of a genome into chromosomes (sometimes called pseudo-chromosomes or more appropriately chromosomal scaffolds), there is scope for other strategies to be employed, creating genome assemblies that are not purely *de novo* due to their reliance on an outside resource. For instance, an existing assembly can be used to inform the assembly of related species (Schneeberger et al. 2011). The obvious trade-off to this approach is the uncertainty it introduces, as even closely related species often have significant differences in genome

structure (Card et al. 2014). The use of linkage maps is also a potential strategy, by using the sequence based markers in the linkage map to anchor matching sequence and contigs to a known map position and building further from there (International Barley Genome Sequencing Consortium 2012; Ren et al. 2012; Kawakami et al. 2014; Bartholomé et al. 2015). This strategy has the advantage of independence from other organisms, and will likely be representative of the true genome architecture of the organism. However, the requirement of creating a mapping cross to build a sufficiently dense linkage map is not an insignificant one. Sequence will also only anchor to areas of the genome where segregation has occurred in the cross and been captured by a marker, which could be biased towards features such as gene rich areas, depending on the marker technology employed (Heslot et al. 2013). In addition, errors in marker ordering occurring as part of linkage analysis will be reflected in the assembly, although it should be noted a statistically sound linkage map can be employed to both check (Myburg et al. 2014) and improve the quality of an existing assembly (Bartholomé et al. 2015). A genome assembly is an evolving resource, and as errors are discovered and new sequencing performed a seemingly 'complete' assembly will be constantly revised. However, even an incomplete non-*de novo* genome assembly opens many avenues of potential investigation.

Applications of genome assemblies

The construction of a genome assembly, while a promising start, will still require the cataloguing of various genomic features to reach its full potential as a resource, including regulatory elements, self-propagating sequence, pseudo-genes and protein coding genes (ENCODE Project Consortium 2012). While manual identification and curation of these features by experts is the most accurate method, the sheer quantity of data being created means that this is usually unfeasible. However, previous study into specific motifs related to genes and gene sequences has permitted the training of computer models to automatically predict the location of protein-coding genes based on sequence motifs observed in other proteins (Meyer et al. 2003; Aziz et al. 2008). All sequences are evaluated for features such as introns, exons and start/stop motifs that make up a coherent functional gene (or other product such as non-coding RNA or transposable element), while also predicting the potential function of the protein product based on the presence of conserved domains (Aziz et al. 2008). Automatic gene annotations such as these have their limitations; for instance the presence of sequencing errors such as frame shifts or insertions or deletions (indels) will have an impact on the models created. A problem unique to eukaryote genomes are the presence of non-coding regions within genes (introns), which can be difficult to predict accurately, especially considering instances of alternate splicing whereby different sections

of the coding region of a gene (exon) may be included or excluded when the mRNA is transcribed (Florea et al. 2005). Often an outside resource such as an expressed sequence tag (EST) library created for the organism in question is required to help with the prediction of these more complex characteristics, such as intron-exon boundaries and 5' and 3' untranslated regions (Stanke et al. 2006). Automatic annotations also have issues with very short or closely spaced genes, often combining multiple separate genes into a single model, especially in the case of tandemly duplicated gene families (Fawal et al. 2014). As the number and genomic location of these gene families is important to the discussion of evolution, manual annotation is often required to ensure these genes are correctly identified. Indeed, manual annotation is the best way to create a highly reliable database, but as long as the aforementioned limitations are accounted for, automatic gene annotations represent a fast and accessible method of evaluating the total gene content of an organism.

When the gene content of a genome has been annotated, there are many avenues of exploration available. Examining the proliferation of specific gene families within an organism and comparative analysis to other related species can reveal much about their evolutionary history, including past and present selective pressures (Cannon et al. 2004). A complete gene catalogue can also be used to test hypotheses relating the phenotype of the organism to gains or losses in biochemical pathways or other gene differences (Veeckman et al. 2016). The locations of QTL are important to investigate in these cases, as genes located within the bounds of a QTL can be flagged as candidate genes influencing the phenotype of interest (Pflieger et al. 2001). However, depending on the QTL confidence interval and the density of genes, there may be many positional candidates for such a QTL, and further investigation methods are often employed, such as examining gene expression (Schunkert et al. 2011). By sequencing the mRNA of individuals experiencing a phenotype of interest (along with control individuals) and mapping these mRNA reads back to the genome, one can examine which genes are active and their level of expression relative to others (Schunkert et al. 2011; Zou et al. 2012). Expression data can also be used to validate gene models (Cantarel et al. 2008), and examine the architecture of differential expression between and within organisms (Ghaemmaghami et al. 2003; Anders and Huber 2010).

Apart from the practical implications of gene annotation, features of the genome can be used to investigate the genomic architecture of the species and its evolutionary history. Aligning the sequence of a genome with one of a closely related species can reveal both areas of high synteny, and areas of differentiation between them, such as chromosome translocations, duplications, deletions or inversions (Parkin et al. 2005). Finding areas that

are heavily repeated, present in several places or absent in other organisms can also be used to infer specific events in the evolutionary history. For instance, analysis of the duplicated genes in the *Eucalyptus grandis* reference genome provided evidence for a whole genome duplication (dated approximately 110 million years ago [MYA]), which was suggested to have been pivotal in the evolution of the Myrtales and diversification from other Rosids (Myburg et al. 2014). Comparative analysis of sequence has also led to the observation that chromosome 3 has undergone almost no inter-chromosomal rearrangement since diverging from its ancestral eudicot ancestor, a situation also observed in another forest tree, *Populus trichocarpa* (Myburg et al. 2014). Insights such as these are only made possible through a well-assembled genome sequence, making these resources invaluable.

In summary, advances in marker and sequencing technology along with more powerful computing and analytical resources have resulted in new possibilities for discovery in the genomics space. It is now feasible with sufficient resources to take a previously unstudied organism to genome assembly without much prior knowledge of the genome, and this 'post-genome' age has led to an abundance of data being generated and a myriad of potential insights to be gained.

Thesis outline

This thesis reports the development and application of genomic tools to the analysis of genetic and genomic architecture in two divergent eucalypt genera: *Eucalyptus* and *Corymbia*. Linkage maps were created and used (alongside pre-existing maps and the *E. grandis* reference genome) to examine differences in genomic architecture between eucalypt species (Chapter 2), and to contrast the genetic control of resistance to an exotic pathogen (*Austropuccinia psidii*) to that of several native pathogens (including *Quambalaria pitereka* and *Teratosphaeria* spp.) via QTL analysis (Chapter 3 & 4). These linkage maps were also used to inform the assembly of the *Corymbia citriodora* subsp. *variegata* reference genome, permitting the annotation of the terpene synthase gene family in *Corymbia* and its subsequent comparison to that of *Eucalyptus* (Chapter 5).

The four experimental chapters are presented as self-contained units in the style of scientific journal articles. Each chapter contains an introduction to the relevant literature, and a discussion of the findings in relation to current knowledge, indicating where advances have been made. Due to this level of discussion in each experimental chapter, Chapter 6 presents a brief general discussion integrating the major findings and the implications of these findings when taken as a whole.

Chapter 2 - Comparative genomics of *Eucalyptus* and *Corymbia* reveals low rates of genome structural rearrangement

Introduction

Comparative genomics is a rapidly expanding field of research, with the potential to provide important evolutionary insights, as well as useful practical information (Arabidopsis Genome Initiative 2000; Drosophila 12 Genomes Consortium 2007; Jaillon et al. 2007; Koonin and Wolf 2008). For example, understanding the genomic similarities and differences between taxa is a central goal of evolutionary genetics, while the identification of conserved genome structure is important for inferring shared ancestry between taxa, and for the transfer of genetic information (Tang et al. 2008). The increasing availability of genomic resources, such as genome sequences and high throughput molecular markers, now provides the opportunity for comparative genomics studies across an ever growing variety of taxa, yielding novel insights regarding the evolution of individual genes or gene families (Carretero-Paulet et al. 2010; Huang et al. 2010; Martin et al. 2010) through to entire genomes (Salse 2012; Tomato Genome Consortium 2012; Nystedt et al. 2013). Species of economic importance such as grasses have been well studied in this regard, while trees have been relatively poorly studied.

Linkage maps are invaluable for the study of genome-wide structural variation between species, especially in the absence of an assembled genome (Kukekova et al. 2007). Linkage maps are a genomic resource that have broad utility, including: the study of quantitative traits (Junghans et al. 2003a; Freeman et al. 2013); comparative genomics (Ahn and Tanksley 1993); analysis of recombination rate (Sakamoto et al. 2000; Groenen et al. 2009); and sequence assembly (International Barley Genome Sequencing Consortium 2012; Kawakami et al. 2014; Bartholomé et al. 2015). Genome structure comparisons can be performed by comparing several maps (Ahn and Tanksley 1993; Shepherd et al. 2006), or comparing maps with assembled genomes (Aitken et al. 2014; Fishman et al. 2014; Wang et al. 2015).

Eucalypts are a group of trees belonging to the Myrtaceae family, containing the genera *Angophora, Corymbia* and *Eucalyptus* (Slee et al. 2006). There are over 700 different species of eucalypts spanning 10 subgenera of *Eucalyptus* and two subgenera of *Corymbia* (Brooker 2000). Most species belong to the *Eucalyptus* subgenus *Symphyomyrtus*, including

many of economic importance such as *Eucalyptus grandis*, *E. urophylla* and *E. globulus* (Doughty 2000; Grattapaglia et al. 2012). *Eucalyptus grandis* is the reference genome for eucalypts (Myburg et al. 2014). Analysis of this genome provided evidence for a whole genome duplication (dated approximately 110 million years ago [MYA]) in eucalypts, which was suggested to have been pivotal in the evolution of the Myrtales and diversification from other Rosids (Myburg et al. 2014). The potential for further genomic studies in these important genera has been greatly enhanced by the release of this resource (Strauss and Myburg 2015). However, the efficacy of information transfer from this reference genome to other species will depend upon their similarity in genome structure, in terms of both synteny (the location of loci on homologous linkage groups) and collinearity (the congruent ordering of loci on homologous linkage groups). Early linkage mapping in eucalypts has allowed comparison of genome structure between *E. grandis* and other symphyomyrts such as *E. urophylla* (Brondani et al. 2006), *E. globulus* (Myburg et al. 2003), as well as *Corymbia* species (Shepherd et al. 2006), with each study reporting no strong evidence for structural differences. However, the relatively small number of markers used for map construction in these studies (such as SSRs and AFLPs) and the need for common markers between maps restricted the resolution of the comparisons that could be drawn. The development of high-throughput, sequence anchored markers in eucalypts has removed these limitations, allowing for much higher resolution genetic maps to be produced and the comparison of linkage maps directly to the reference genome (Kullan et al. 2011; Neves et al. 2011; Petroli et al. 2012; Bartholomé et al. 2015; Li et al. 2015; Silva-Junior and Grattapaglia 2015). One recent study using such markers found support for two small inter-chromosomal translocations between *E. globulus* and *E. grandis* x *E. urophylla* hybrids (Hudson et al. 2012b); one of which was supported by replication (independently constructed linkage maps) making it the most definitive genomic difference discovered in eucalypts. Aside from this, a high degree of genome conservation was assumed between members of *Symphyomyrtus* based on all past studies (Grattapaglia et al. 2012). Only one study has performed comparisons outside of this subgenus into the more distant *Corymbia* (Shepherd et al. 2006), but was limited in the number of shared markers. With the advancement of marker technologies more comprehensive comparisons can be made between more divergent eucalypt taxa.

Corymbia, only recently classified as a separate genus to *Eucalyptus* (Hill and Johnson 1995), includes 113 species (Parra-O. et al. 2009), with most endemic to the tropics, arid, and semi-arid zones of northern Australia (Hill and Johnson 1995). Of these, *Corymbia citriodora* subsp. *variegata* (spotted gum) is a species with a prominent role in forestry both in Australia and overseas (Rockwood et al. 2008), where it is used for products including

timber, charcoal and essential oil (Asante et al. 2001; Lee 2007; Rockwood et al. 2008). *Corymbia citriodora* subsp. *variegata* can readily hybridize with *C. torelliana,* an invasive tree species (Wallace and Trueman 1995; Wallace et al. 2008) from the same subgenus but a different section of *Corymbia* (Hill and Johnson 1995; Parra-O. et al. 2009). *Corymbia torelliana* is of interest to forestry due to the potential for increased growth rate in hybrids (Lee et al. 2009; Dickinson et al. 2010). *Corymbia citriodora* subsp. *variegata* and *C. torelliana* have estimated genome sizes of 370 MB and 390 MB, respectively (Grattapaglia and Bradshaw Jr 1994), which is in contrast to the much larger *E. grandis* genome of 640 MB (Grattapaglia and Bradshaw Jr 1994). Despite these differences in genome size, both *Corymbia* and *Eucalyptus* share the same chromosome number, which is conserved across all eucalypts (Grattapaglia and Bradshaw Jr 1994) and indeed across most Myrtaceous species (Grattapaglia et al. 2012). *Corymbia* and *Eucalyptus* separated an estimated 52 MYA (Crisp et al. 2011; Thornhill et al. 2015), and the extent to which changes in genome structure have accumulated in that time and contributed to differences in genome size are unknown.

The extent of genomic differentiation between taxa, and the rate at which this accumulates, has important practical and evolutionary implications. These include influencing reproductive isolation as well as recombination in interspecific hybrids (Stebbins 1950; Dvořák and Zhang 1992). There is increasing evidence that woody perennials are characterised by relatively slow rates of genomic change, whether at the level of substitution rate, chromosomal structure or ploidy (Jaillon et al. 2007; Lanfear et al. 2013; Luo et al. 2015). For instance, a cytological study comparing various woody genera within the Fagaceae family found varying genome size, but no instances of polyploidy contributing to the diversification of this family (Chen et al. 2014). Likewise, a comparative genomic study found a high amount of structural conservation between northern hemisphere trees from genera *Vitis, Populus, Malus* and *Juglans,* relative to herbaceous genera such as *Arabidopsis* and *Medicago* (Luo et al. 2015). Indeed, comparisons between the genomes of herbs and grasses often reveal highly divergent structure, with studies detailing high chromosome fragmentation and ploidy changes (Blanc et al. 2000; Swigonova et al. 2004; Parkin et al. 2005; Tennessen et al. 2014). However, high resolution comparative genomics studies have largely been restricted to a few tree families, such as Fagaceae (Bodénès et al. 2012), Pinaceae (Sakaguchi et al. 2015) and Salicaceae (Hou et al. 2016), therefore it is yet to be seen whether a reduced rate of genomic change compared to herbs is a characteristic of most trees.

In this study we compare the genome structure of the eucalypt genus *Corymbia* to that of *Eucalyptus*. Using 15,360 sequence-based Diversity Array Technology (DArTseq) markers and a marker binning technique (Sun et al. 2007; Jighly et al. 2015) we created high density linkage maps for *C. citriodora* subsp. *variegata* (CCV) and *C. torelliana* (CT). These maps were used to compare genome structure between each parental species and between these *Corymbia* species and *E. grandis* using the reference genome. We present evidence for differences in genome structure which are discussed in the context of the evolutionary relationships between the species and the stability of plant genomes through evolutionary time.

Material and methods

Genetic material

Three genetic linkage maps were generated using two *Corymbia citriodora* subsp. *variegata* (CCV) x *Corymbia torelliana* (CT) F_1 hybrid pedigrees (360 seedlings), resulting from a cross of the same CCV pollen parent (1CCV2-054) with two different CT parents (1CT2-018 and 1CT2-050, Figure 2.1).

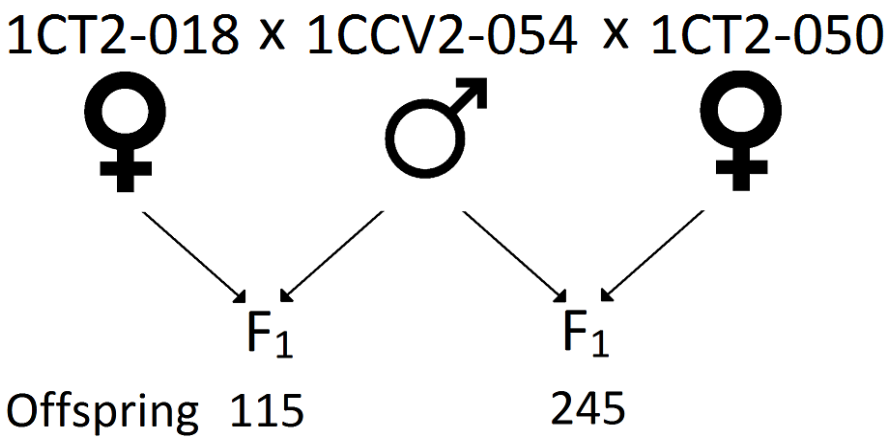


Fig. 2.1 Design of the *Corymbia* pedigrees used to create the linkage maps. CT refers to *Corymbia torelliana*, while CCV refers to *Corymbia citriodora* subsp. *variegata*.

DNA extraction protocol

Offspring were grown in glasshouse conditions until approximately 50 cm tall before sampling. Leaf samples were taken from each individual in the mapping family (including the parents) and dried over silica gel prior to DNA extraction. Total genomic DNA was extracted from 100 mg of dry leaf tissue using a QIAGEN (Hilden, Germany) DNeasy Plant Maxi Kit. The standard protocol was modified as follows: the volume of the AP1 buffer was increased to 1.5 x standard (i.e. 600 µl), 2% PVP-40 was added to the tissue lysis solution, and DNA was loaded onto the spin columns over two centrifugations before elution to

increase yield. DNA samples were concentrated by vacuum drying and quantified using a PicoGreen assay (Molecular Probes, Eugene, OR). Samples were then adjusted to achieve a target concentration of 50 ng/μl by either dilution in 1X TE buffer, or further concentration using a sodium acetate precipitation, where appropriate. 15 μl of solution was supplied to DArT.

DArTseq genotyping

Genotyping was performed by Diversity Array Technology Pty. Ltd. (Canberra, Australia) using DArTseq technology (Sansaloni et al. 2011), which generates 64 base pairs (bp) of sequence at each marker by next generation sequencing. DArTseq yields two types of markers based on sequencing of genomic representations; co-dominant single nucleotide polymorphisms (SNP) and dominant markers which may represent SNP, or length polymorphisms in restriction enzyme recognition sites or restriction fragments. All test-cross (i.e. uniparentally segregating) markers were recoded into a double haploid configuration to allow a marker binning process to take place as SIMPLEMAP (Jighly et al. 2015) requires population data in this format (see the Linkage map construction section). These markers were grouped into quality classes for the different mapping approaches using the following parameters (supplied by DArT PL.): reproducibility; call rate; and polymorphism information content (PIC). The latter is a measure of segregation ratios (a PIC of 0.5 indicates perfect 1:1 segregation) (Shete et al. 2000). First class dominant markers featured reproducibility 1.0, call rate > 95%, and PIC > 0.35 (SNP markers featured average PIC > 0.20); the second class featured reproducibility > 0.9, call rate > 90% and PIC > 0.25 (SNP markers average PIC > 0.15). A third class of markers, used only within bins, featured reproducibility > 0.9, call rate > 80% and PIC > 0.15 (SNP markers average PIC > 0.10, or markers with ambiguous or impossible segregation data, which was resolved by correcting offspring genotypes based on the segregation of the parents [assessed from three replicates of each parent]). Markers that did not meet these thresholds were excluded from further analysis. Fully informative inter-cross markers that segregated 1:1:1:1 were recoded into separate loci, each displaying the alleles segregating from a single parent (i.e. into the double haploid configuration required by SIMPLEMAP).

Linkage map construction

The vast number of molecular markers provided by high-throughput technologies, such as DArTseq, challenges conventional approaches for linkage mapping. A marker binning process in SIMPLEMAP (Jighly et al. 2015) was used prior to map construction to increase computational efficiency and improve the accuracy of high-density map construction

(Collard et al. 2009). In the binning process, a single representative marker is identified which represents a set of co-segregating and tightly linked markers in each bin (hereafter referred to as a 'bin marker'). The bins are created according to a user defined maximum number of recombination events (the 'repulsion threshold') between any pair of markers. SIMPLEMAP recommends a maximum repulsion threshold equivalent to 3 cM for the Kosambi mapping function (Jighly et al. 2015), which for small map distances is equivalent to a recombinant frequency of approximately 3% (Sturtevant 1913). Therefore, a repulsion threshold of three and seven recombinants was used for cross 1CT2-018 x 1CCV2-054 and cross 1CT2-050 x 1CCV2-054 respectively, equivalent to a recombination frequency of less than less than 3% in each cross.

All mapping was undertaken using JoinMap v4 (Van Ooijen 2006). In summary, individual parental maps were initially constructed for both pedigrees using only bin markers (hereafter termed 'bin maps'). These maps were used to assess biological and technical replication between the parental maps. Comprehensive parental maps were then constructed for each pedigree, using all markers to provide a higher resolution comparison against the *E. grandis* genome. A summary of the methods is presented below (Figure 2.2).

Parental bin map construction

Separate bin maps were created for both parents in each cross using only first class markers segregating 1:1. After removing markers and individuals with > 10% missing data, markers were placed into linkage groups at a minimum of LOD 3. The regression algorithm and Kosambi mapping function (Kosambi 1943) were used to order markers within linkage groups, using default JoinMap v4 settings. In an attempt to construct maps with robust marker order, an iterative approach was used and stringent criteria were imposed to evaluate map orders and remove problematic markers in each linkage group. Specifically, markers with a Chi-square goodness-of-fit contribution > 1.0, or present in > 1 double crossover were excluded. Markers with segregation distortion widely different from their closely-linked markers, were also excluded as these were likely to represent genotyping errors (Van Ooijen 2006). After removing markers according to these criteria, linkage maps were re-calculated and the above criteria were again evaluated. This procedure was repeated until threshold values were reached by all markers in each linkage group. Maps were then recreated using the Maximum Likelihood algorithm and compared with those created by the regression algorithm to verify marker order. To avoid interpreting potential error in the ordering of tightly linked markers as a departure from collinearity of syntenic markers (see the Discussion section), a threshold of 1 cM was used to detect non-collinearity. Any shift in marker position exceeding this threshold between the maps was

criteria for re-evaluation of marker statistics and further removal based on statistical support for marker order, until collinearity was established between maps produced using the different algorithms.

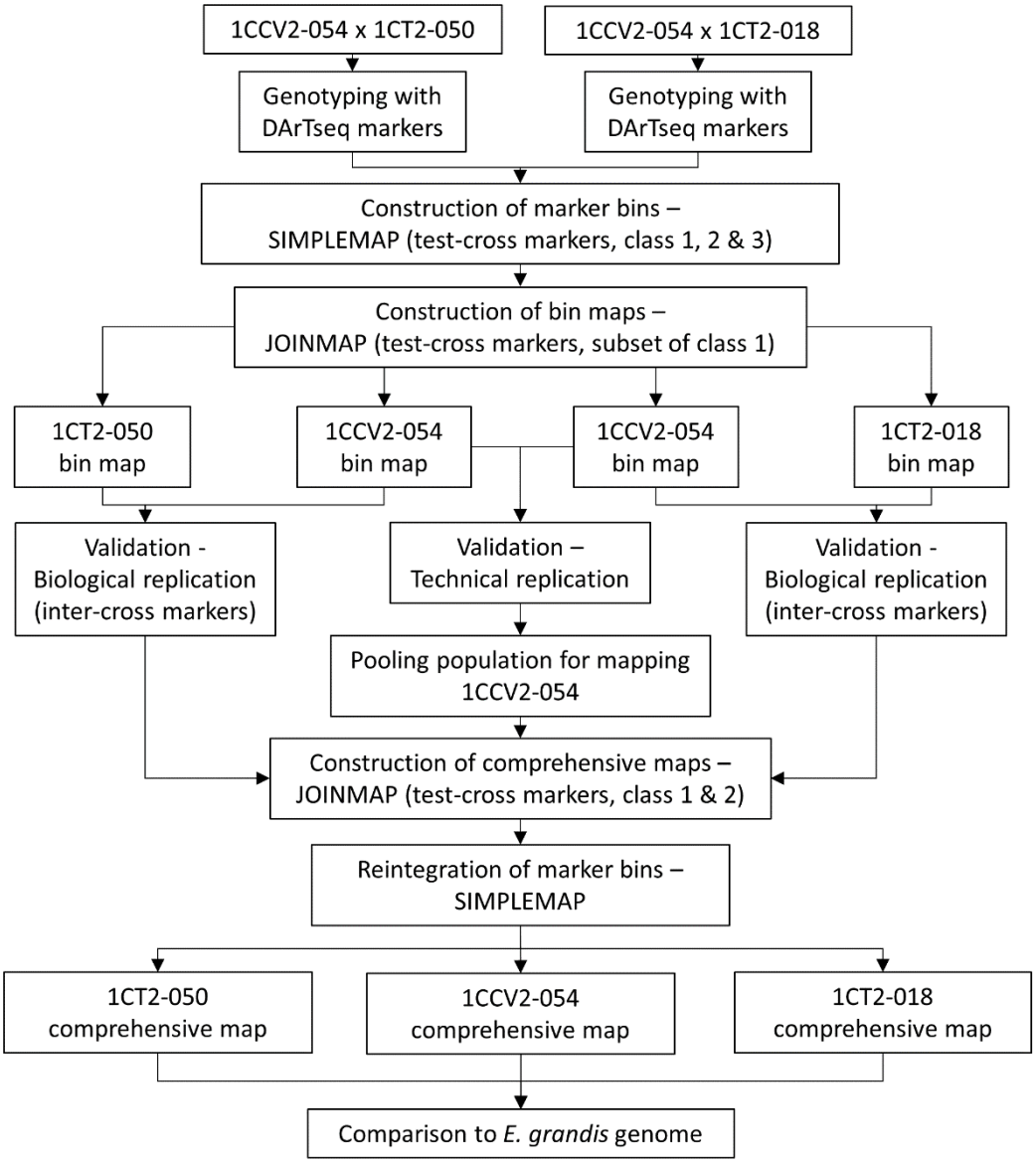


Fig. 2.2 Summary of methods followed to create *Corymbia citriodora* subsp. *variegata* and *Corymbia torelliana* linkage maps.

Comparison of parental bin maps

In order to evaluate the repeatability of marker ordering, both biological and technical replication was evaluated by calculating Spearman's correlations between linkage groups in different maps. Technical replication was evaluated by comparing the order of markers in each linkage group between the two independent maps of the male parent (1CCV2-054) which is shared between crosses. The technical replicates represent meiosis from the same genotype sampled in two different crosses. Biological replication was evaluated by comparing CCV and CT maps within crosses, as they represent different samples of meiosis

and are from different genetic material (i.e. different species of *Corymbia*). The latter required mapping inter-cross (i.e. bi-parentally segregating, 1:2:1 and dominant 3:1) markers together with the test-cross (1:1) bin markers, to allow direct comparison of the parents within each pedigree based on common markers. When adding inter-cross markers, the ordering of markers in each round of mapping was evaluated as described above for bin maps, with removal biased towards retaining these bi-parentally segregating markers.

Mapping of CCV using segregation data from the two different populations, and comprehensive map construction

Given the high correlation between CCV bin maps (see results below), genotype data from both populations was combined to map test-cross markers segregating from the CCV parent in both crosses, and this dataset was treated as a single population (n = 360). Due to the increase in sample size, markers were re-binned with SIMPLEMAP using a repulsion threshold of 10 recombinants to create bins spanning less than 3 cM.

Mapping of this combined CCV dataset (and the original CT datasets) was undertaken using only first and second class bin markers. After removing markers and individuals with > 10% missing data, markers were added in an iterative fashion, starting with approximately 500 high quality markers, with batches of around 250 lower quality markers added in subsequent rounds. Markers were grouped at a minimum of LOD 3, and were ordered using the regression algorithm and Kosambi mapping function (Kosambi 1943). In each iteration, markers were removed according to the criteria described above, except the threshold for Chi-square goodness-of-fit contribution was raised to > 2 after the initial 500 markers. Marker order was verified by comparing the final map from the previous round to the map produced using the Maximum Likelihood algorithm. Any shift in marker position exceeding 1 cM in any of these comparisons was criteria for re-evaluation of marker statistics and removal according to statistical support for marker order as above (in this case also considering the quality ranking of each marker), until collinearity was established.

Subsequently, ‘comprehensive maps’ were constructed, in which all markers (including first, second and third class) from the binning procedure were reintegrated into the bin maps. In SIMPLEMAP, reintegration of binned markers is performed using the percentage of recombinants between two markers to order the markers within bins around their representative bin marker, whose position is fixed (Jighly et al. 2015).

Comparison with Eucalyptus grandis

To compare the genome architecture of the *Corymbia* species and *E. grandis*, the marker sequences (length 64 bp) were searched against the *E. grandis* genome v2 (Myburg et al. 2014; Bartholomé et al. 2015) to identify putative sequence homologs, using BLASTN (Altschul et al. 1997). For each marker the highest scoring hit (if multiple) was accepted only if it exceeded 95% of query coverage, and had an e-value < $1e^{-10}$. Markers that fell on unanchored *E. grandis* scaffolds were not considered. In order to examine synteny (the location of loci on homologous linkage groups) and collinearity (the congruent ordering of loci on homologous linkage groups) between *Corymbia* and *E. grandis*, the physical position of these hits in *E. grandis* were plotted against genetic position on the CCV map, and marker order was compared using Spearman's rank correlation. Given the high collinearity of syntenic markers discovered between *E. grandis* and *Corymbia*, linkage group numbering and the orientation of linkage groups for the CCV and CT maps followed Brondani et al. (2006), which corresponds to the chromosomes of the *E. grandis* reference genome (Myburg et al. 2014).

To determine if there were any detectable instances of inter-chromosomal duplication involving multiple collinear markers in *E. grandis* relative to CCV, a second round of BLAST was undertaken allowing for multiple high scoring pairs per marker, and the position of these hits was compared to the CCV map as above.

Results

DArTseq genotyping

After preliminary data analysis to remove poor quality markers, DArTseq genotyping yielded 10,726 markers segregating 1:1 from the 1CCV2-054 individual, and 6,554 and 6,323 segregating 1:1 from 1CT2-050 and 1CT2-018, respectively, across the three quality classes described above (Supp. 2.1). Dominant markers made up the bulk of the total, with co-dominant SNP markers averaging 25% of the markers across each individual.

Comparison of parental bin maps

The bin maps for each parent in the two crosses comprised 340 to 446 bin markers. The rank order of the two bin maps of 1CCV2-054 (technical replicates) were highly correlated, providing strong support for the marker order (Table 2.1) and for the approach of combining the two populations to produce a comprehensive map of 1CCV2-054. Likewise, the high rank order correlation of the parental maps within pedigrees (biological replicates) provided good support for map order, and implied the genomes of the two *Corymbia*

species are highly collinear. However, correlations could not be carried out for linkage group 5 and 11 (Table 2.1) due to insufficient bi-parentally segregating markers.

Table 2.1 Spearman's correlation of marker order in *Corymbia citriodora* subsp. *variegata* and *Corymbia torelliana* bin maps.

Linkage Group	Spearman's correlation[a]					
	1CCV2-054 vs 1CCV2-054 (bin markers)		1CT2-050 vs 1CCV2-054 (inter-cross markers)		1CT2-018 vs 1CCV2-054 (inter-cross markers)	
1	1.00***	(25)	0.80	(4)	0.93***	(9)
2	1.00***	(27)	1.00***	(3)	1.00***	(9)
3	0.99***	(19)	1.00**	(5)	0.96***	(14)
4	1.00***	(32)	1.00***	(4)	0.98***	(13)
5	1.00***	(31)	NA	(1)	1.00***	(7)
6	1.00***	(23)	0.99***	(17)	1.00***	(4)
7	0.99***	(21)	1.00**	(6)	1.00***	(4)
8	1.00***	(34)	1.00***	(7)	0.97***	(9)
9	1.00***	(28)	0.60	(5)	0.90*	(5)
10	1.00***	(23)	1.00***	(7)	1.00***	(11)
11	1.00***	(23)	0.94*	(6)	NA	(2)

[a]Numbers in brackets indicate the number of shared markers present between each bin map, while the type of marker is specified in the column heading. NA indicates linkage groups where less than three common inter-cross markers (dominant markers segregating 3:1 and SNP markers segregating 1:2:1) were able to be ordered, so no correlation was possible. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$.

Comprehensive maps

The number of markers in the comprehensive maps ranged from 4,616 – 6,055, while map length ranged from 1,115 – 1,346 cM (Table 2.2, Supp. 2.2). Marker density was high, with mean marker interval ranging from 0.26 – 0.61 cM. The map constructed for 1CT2-018 had the greatest length, mean and maximum marker interval, likely due to the relatively small population size used for map construction, as shown in a simulation by Bartholomé et al. (2015). The technical replicates created for 1CCV-054 support this observation, with the bin map created in the smaller pedigree also displaying a greater length, mean and maximum marker interval compared to the bin map from the larger pedigree (results not shown).

Table 2.2 Description of the comprehensive linkage maps generated for *Corymbia citriodora* subsp. *variegata* and *Corymbia torelliana.*

Mapped individual		Population size	Length (cM)	Linkage group length (cM)	Markers	Unique positions	Mean interval between markers (cM)[a]	Maximum interval (cM)
1CCV2-054	♂	360	1,179.9	77.2 - 137.6	6,055	4,510	0.26	10.5
1CT2-050	♀	245	1,114.8	79.0 - 126.4	4,689	2,834	0.39	9.3
1CT2-018	♀	115	1,345.6	94.8 - 158.7	4,616	2,212	0.61	15.5

[a]Mean marker interval was calculated from unique positions.

Collinearity of Corymbia citriodora subsp. variegata with Eucalyptus grandis

Of the 6,055 markers ordered on the CCV comprehensive map, 1,441 were matched to a position on the *E. grandis* genome (Myburg et al. 2014) at the threshold for acceptance (> 95% query length, e-value < $1e^{-10}$, highest scoring hit) (Figure 2.3). Additionally, 204 CCV markers mapped to minor *E. grandis* scaffolds. Of the markers anchored to one of the 11 chromosomes, 165 (11%) were non-syntenic and 320 (22%) of the syntenic markers were non-collinear. Only markers that were at least 2 MB removed from the collinear order were declared as non-collinear to avoid interpreting possible error associated with ordering tightly linked markers as non-collinearity (see the Discussion section). There was a significant positive correlation in the order of syntenic markers between CCV and *E. grandis* (Table 2.3). Of the 1,441 markers which were placed on *E. grandis* chromosomes, 449 had more than one identically scored hit on the same chromosome, but as the majority of these were within 2 MB of each other and would have no impact on collinear order, one was selected at random. These multiple hits potentially reflect the numerous duplicate genes in tandem arrays known to be present in *E. grandis* (Myburg et al. 2014).

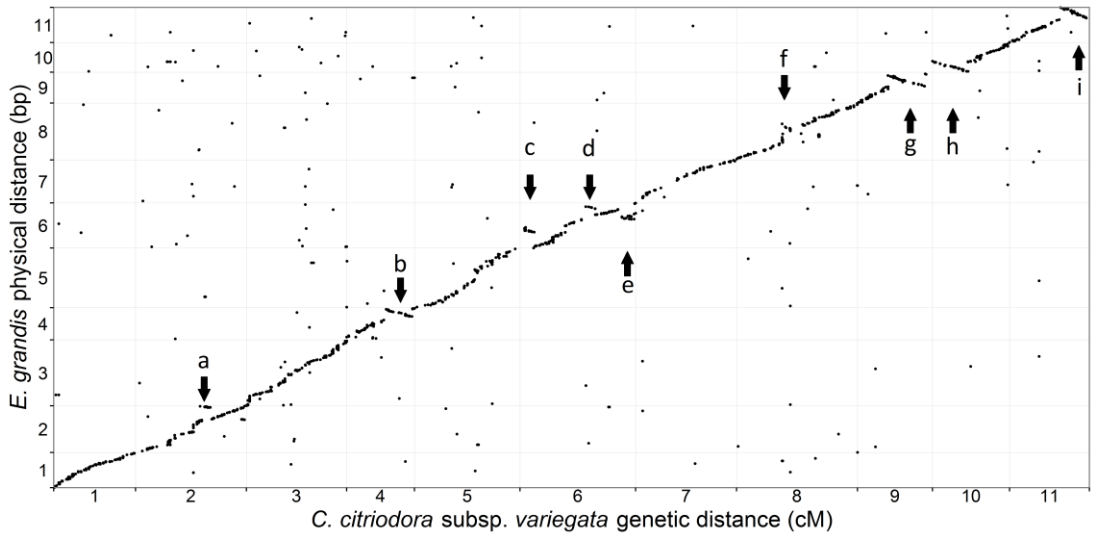


Fig. 2.3 Marker positions in the *Corymbia citriodora* subsp. *variegata* comprehensive linkage map relative to the *Eucalyptus grandis* genome. Numbers along the x and y axis indicate the chromosome boundaries. Terminal inversions were detected in *C. citriodora* subsp. *variegata* relative to *E. grandis* on chromosomes 4, 9, 10 and 11; an intra-chromosomal translocation on chromosome 2; and more complex rearrangements on chromosome 6 and 8. The position of the above rearrangements are indicated by arrows, and named following Table 2.4. This figure was created using the package 'ggplot2' (Hadley 2009) in R (R Core Team 2017).

Table 2.3 Marker order correlation between the *Corymbia citriodora* subsp. *variegata* map and the *Eucalyptus grandis* genome.

Linkage Group	Spearman's correlation[a]
1	0.99***
2	0.84***
3	0.95***
4	0.87***
5	0.98***
6	0.77***
7	0.98***
8	0.97***
9	0.71***
10	0.92***
11	0.79***

[a]Correlations below 0.95 are found on those chromosomes where putative rearrangements were found. *** $P < 0.001$

The analysis of collinearity provided evidence for nine major chromosomal rearrangements (involving consecutive non-collinear markers spanning > 5cM) between *E. grandis* and CCV, occurring on seven linkage groups. Specifically, large terminal inversions were evident on linkage groups 4, 9, 10 and 11, and more complex rearrangements detected on linkage groups 2, 6 and 8 (Table 2.4). To provide support for these putative rearrangements both of the maps generated for the CT parents were compared to the *E. grandis* reference genome (Supp. 2.3). Of the nine described rearrangements, seven were also present in both CT maps, while the rearrangements on linkage groups 2 and 8 could not be validated due to low marker density in these areas of the CT maps.

Table 2.4 Position of putative rearrangements in *Eucalyptus grandis* relative to *Corymbia citriodora* subsp. *variegata*.

Chromosome[a]	Name[b]	Type	Position of markers flanking rearrangement (bp)[c]		Position of markers spanning rearrangement (bp)[d]	
2 (a)	CCV-in(2)tp1	Inversion/ translocation	41,313,232	42,473,364	57,510,947	58,000,349
4 (b)	CCV-in(4)1	Inversion	25,483,501	40,126,737	29983251	38,988,874
6 (c)	CCV-in(6)tp1	Inversion/ translocation	Start	2,162,368	20,167,425	24,639,366
6 (d)	CCV-in(6)tp2	Inversion/ translocation	35,523,433	42,332,855	50,159,071	52,725,864
6 (e)	CCV-in(6)tp3	Inversion/ translocation	48,934,608	56,404,168	36,780,165	39,438,890
8 (f)	CCV-in(8)1	Inversion	22,156,661	45,300,454	33,441,444	42,114,915
9 (g)	CCV-in(9)1	Inversion	20,700,201	37,191,595	21,585,756	33,994,985
10 (h)	CCV-in(10)1	Inversion	Start	13,232,993	1,214,529	14,051,146
11 (i)	CCV-in(11)1	Inversion	29,255,951	End	31,407,473	44,623,976

[a]The letter assignment corresponds to the naming of the rearrangement in Figure 2.3.
[b]Designation of each inversion. 'in' refers to an inversion, the number in brackets refers to the linkage group the rearrangement is localized to, 'tp' indicates the rearrangement is transposed within the chromosome, and the final number indicates occurrence on the chromosome, if multiple.
[c]Refers to first marker on either side of the rearrangement. No flanking marker position was available if a rearrangement spanned the first or last marker on a linkage group in the *C. citriodora* subsp. *variegata* comprehensive map.
[d]Refers to the marker in the first and last position of the rearrangement. Note, for translocations the position of markers flanking a rearrangement indicates the origin of the translocated region in *E. grandis*, while the position of markers spanning the rearrangement indicates the current configuration in the *C. citriodora* subsp. *variegata* comprehensive map.

To investigate the possibility that the putative rearrangements were artefacts caused by errors in the *E. grandis* genome assembly, the areas of the *E. grandis* genome containing the nine putative rearrangements were checked to ensure collinearity with independently constructed high density linkage maps constructed in *E. grandis* and *E. urophylla* (Bartholomé et al. 2015). For this purpose, the physical location in *E. grandis* of all putative rearrangements (including two markers flanking the rearrangement) was assessed for collinearity with the genetic linkage maps (J. Bartholomé, pers. comm.). These areas of the *E. grandis* genome were highly correlated with the marker order in both the *E. grandis* and *E. urophylla* linkage maps (Supp. 2.4), giving confidence that these areas of the genome were assembled correctly.

The nine major intra-chromosomal rearrangements described above involved 200 (14%) of the 1,441 CCV markers placed on the *E. grandis* chromosomes. The remaining 120 (8%) non-collinear markers detected in this comparison were mostly singletons, but also included small clusters of tightly linked markers none of which spanned more than 5 cM (Figure 2.3; Supp. 2.2). Similarly, while distributed genome-wide, the majority of the 165 non-syntenic markers also occurred as singletons or in near identical positions to other markers, with no consecutive markers spanning more than 5 cM (Figure 2.3; Supp. 2.2), suggesting no major inter-chromosomal rearrangements have occurred between these species. Likewise, no instances of inter-chromosomal duplications involving multiple collinear markers were detected when examining markers with multiple matches on the *E. grandis* genome (Supp. 2.5).

Discussion

We perform the first detailed comparisons of genome structure between *Corymbia* species, as well as between each species and the *Eucalyptus grandis* reference genome. The results of these comparisons provide the first evidence for large scale chromosome rearrangements in eucalypts. Previous comparative studies of eucalypts have pointed to largely conserved genome structure (Grattapaglia et al. 2012; Freeman 2014). However, detailed comparisons have been restricted to a few species within subgenus *Symphyomyrtus* (Kullan et al. 2011; Neves et al. 2011; Hudson et al. 2012b; Li et al. 2015). Comparison of *C. torelliana* and *C. citriodora* subsp. *variegata* linkage maps (both directly and *via* comparison of these linkage maps with the *E. grandis* genome), suggests genome structure is largely conserved between these *Corymbia* species. These species represent separate sections within *Corymbia* (Hill and Johnson 1995), so in terms of taxonomic distance are comparable to the previous inter-sectional comparisons within *Symphyomyrtus* (Myburg et al. 2003; Hudson et al. 2012b; Li et al. 2015). In contrast, much greater genomic differentiation was evident in our comparison between the closely related genera *Eucalyptus* and *Corymbia*. Together with past findings our results provide further evidence that genome structure is highly conserved between closely related eucalypt species with more pronounced genomic differentiation found with increasing taxonomic distance.

Despite rearrangements being detected on seven linkage groups, the genomic structural differentiation found between the two genera in this study is low in the context of many plant taxa, such as *Arabidopsis, Sorghum, Zea, Brassica* and *Fragaria* (Blanc et al. 2000; Swigonova et al. 2004; Parkin et al. 2005; Tennessen et al. 2014), but comparable with the high level of genomic stability reported in other woody angiosperms. For example, while

Salix (willow) and *Populus* (poplar) diverged approximately 45 - 52 MYA, comparative mapping (Hanley et al. 2006; Berlin et al. 2010), and comparison of assembled genomes (Dai et al. 2014) reveal high synteny and collinearity between the two genera. Likewise, *Castanea* (chestnut) and *Quercus* (oak) diverged approximately 70 MYA, but comparative mapping based on 397 shared markers revealed conserved chromosome number and high collinearity (Bodénès et al. 2012). In contrast, grasses and herbaceous plants often display chromosome reshuffling and changes in ploidy level between more recently diverged species (Soltis et al. 2014). Ploidy is stable throughout the Myrtaceae (Grattapaglia et al. 2012) and most other trees, with some exceptions (Wilkinson 1944; Sterck et al. 2005). Although woody angiosperms do not form a single evolutionary lineage, shared characteristics such as their large size and longevity influence their mode and tempo of evolution (Petit and Hampe 2006) and this may extend to genome structure (Bodénès et al. 2012; Hou et al. 2016). Specifically, our findings from a geographically and phylogenetically independent angiosperm lineage from those in previous comparisons support the hypothesis that conservation of genome structure is a key evolutionary characteristic of trees (Chen et al. 2014; Luo et al. 2015).

There are several potential explanations for conservation of genomic structure amongst diverse woody angiosperms. The disparity in the rate of genome structural changes between herbaceous and non-herbaceous plants may simply reflect differences in generational time, with more rapid genomic differentiation occurring in organisms with faster generation turnover relative to woody perennials (Sinnott 1916; Luo et al. 2015). Further, Chen et al. (2014) proposed that participation in syngameons (populations of different species with interbreeding) may play an important role in the conservation of genome structure in woody angiosperms. The premise is that syngameous relationships may promote genome conservation because inter-specific gene flow can be advantageous, potentially allowing rapid adaptation without the need for major genomic changes. Indeed, hybridisation has long been hypothesised to play an important role in eucalypt evolution (Ashton and Sandiford 1988; Griffin et al. 1988; Potts et al. 2003; McKinnon et al. 2004). Hybridisation in eucalypts is more frequent between closely related species and drops off sharply with increasing taxonomic distance (Potts and Dungey 2004; Dickinson et al. 2012; Larcombe et al. 2015). For example, the symphyomyrts *E. grandis*, *E. urophylla* and *E. globulus* can all interbreed, as can the two *Corymbia* species in this study, consistent with the apparent conservation of genome structure between species within each of these genera (Myburg et al. 2003; Hudson et al. 2012b). However, *Eucalyptus* and *Corymbia* do not hybridise with one another (Griffin et al. 1988). Assuming that interspecific hybridisation does contribute to genome conservation in closely related eucalypt species,

one can speculate that the bulk of the putative rearrangements between *Eucalyptus* and *Corymbia* would have been selected against in a syngameous relationship, and may have occurred after these lineages were reproductively isolated. However, further study is required to better understand the evolution of genome structure between these genera and in eucalypts more broadly, ideally performing comparative genomics and phylogenetic analysis of several taxa representing different lineages.

The expansion and contraction of gene families by tandem duplication is another potential factor which may contribute to taxonomic differentiation amongst eucalypts while conserving gross genome structure (Yang et al. 2008; Freeling 2009; Liu et al. 2012). Tandem duplication is thought to be a major mechanism creating new genes with implications for adaptation and speciation (Taylor et al. 2001; Long et al. 2003; Roth et al. 2007). This may be particularly true in eucalypts, as *Eucalyptus grandis* has the largest proportion of genes in tandem repeats among sequenced plant genomes. Indeed, preliminary analysis points to variation in copy number of tandem repeats in comparison with the closely related *Eucalyptus globulus* (Myburg et al. 2014), providing support for the role of tandem duplication in eucalypt diversification.

The use of a marker binning technique, iterative rounds of mapping and stringent thresholds for accepting a given map order contributed to very robust marker orders, as evidenced by the strong correlations between maps in this study. To our knowledge, these are the highest density linkage maps published in eucalypt to date (Bartholomé et al. 2015; Silva-Junior and Grattapaglia 2015). Establishing the correct map position of tightly linked markers in high-density linkage maps is statistically challenging (Hackett and Broadfoot 2003; Collard et al. 2009). To alleviate this problem a marker binning technique was employed, which grouped tightly linked markers into bins before ordering. This was effective in reducing the computational complexity of mapping thousands of markers, and should have reduced gross errors which occur more frequently when attempting to order tightly linked markers (Collard et al. 2009). As genotyping errors and missing data are also key factors producing incorrect marker order, particularly as marker density increases (Hackett and Broadfoot 2003), our iterative approach of progressively increasing marker density from the highest quality markers (which generally contain the least genotyping errors and missing data) to those of lower quality gave us confidence in the marker orders produced and permitted an assessment of repeatability of marker orders.

The creation of individual parental maps based predominantly on test-cross markers also contributes to robust orders. Past studies have often employed bi-parental consensus maps incorporating inter-cross markers (Kullan et al. 2011; Hudson et al. 2012b; Petroli et al.

2012). Such consensus maps have the advantage of allowing comparison of male and female maps and the location of QTL, and can result in increased marker density. However, a consensus map may be less robust, as inter-cross markers have been shown to reduce the accuracy of marker ordering (Bartholomé et al. 2015). Indeed, in this study when attempting to add inter-cross markers to compare the parental maps, only a few could be ordered at the required stringency. Further, merging parental maps can create errors due to heterogeneity between individuals used for map construction (Gustafson et al. 2009). One of the main outcomes we sought to achieve through the creation of these maps was to inform a *Corymbia* genome assembly (Shepherd et al. 2016), along with facilitating further comparative genomics among eucalypts. Therefore, we chose to use individual parental maps with an emphasis on stringent marker order, rather than maximising the number of markers placed on a single map.

Both linkage maps and genome assemblies are prone to errors (Van Ooijen 2006; Collard et al. 2009; Hamilton and Robin Buell 2012; Bartholomé et al. 2015) and should be independently validated where possible, in order to draw robust conclusions in comparative studies. However, studies of this nature rarely have replication. In our case, the majority of the rearrangements (seven out of nine) we describe are supported by independently constructed linkage maps in this study, providing replication; both within *C. citriodora* subsp. *variegata* and in a separate species, *C. torelliana*. The areas of the *E. grandis* genome assembly in which these putative rearrangements lie have also been validated through comparison of the *E. grandis* genome to independently constructed linkage maps (Bartholomé et al. 2015). As such, we are confident these putative rearrangements reflect real genomic differences between the taxa in question, rather than errors in linkage map construction or genome assembly.

Aside from the nine relatively large rearrangements, many smaller regions were non-syntenic or non-collinear in the comparison of the CCV linkage map and the *E. grandis* genome. These regions were dispersed throughout each genome with the majority represented by single markers, but also included small groups of (up to five) markers. The placement of these markers likely represents both small genomic differences and analytical causes. In the case of the latter, despite the use of replication and stringent methodology errors may occur due to factors such as incorrect order (Collard et al. 2009) or linkage group assignment of mapped markers; errors in the *E. grandis* genome assembly (Bartholomé et al. 2015); and failure of BLAST to locate the true *E. grandis* homolog of markers in the CCV map. On the other hand, some differences are likely to reflect biological causes including small scale inversions, duplications and deletions as well as transposable

element activity between the genomes, which have been implicated in inter-chromosomal rearrangements in eucalypts (Hudson et al. 2012b) and other taxa (Feschotte et al. 2002; Morgante et al. 2007). In eucalypts an increasing level of small scale non-synteny was noted when comparing taxa with increasing taxonomic separation (Hudson et al. 2012b), so the level of non-synteny shown between these genera is not unexpected. Despite the fact that some of the apparent small genomic differences no doubt represent errors, overall, the linkage maps created in this study provide valuable insights into the extent of genome differentiation between *E. grandis* and *Corymbia* and highlight potential differences for further research.

Researchers are currently using the *E. grandis* reference genome for gene discovery across many eucalypt species while assuming conservation of genome structure, but our findings show this requires validation, particularly in divergent lineages such as *Corymbi*a. Large stretches of conserved marker orders were found between the genomes of *Corymbia* and *E. grandis*, with even those areas encompassed by putative rearrangements maintaining a conserved order within the inversions. These findings suggest that information regarding broad scale genomic features will be readily transferable between the two genera. However, transfer of information at the genic scale, such as the content and order of annotated genes (Myburg et al. 2014; Külheim et al. 2015; Christie et al. 2016), as well as the potential impact of expansion and contraction of genes in tandem arrays in *Corymbia,* will require further analyses at the sequence level. The putative rearrangements revealed in this study are likely to be of relevance to these analyses.

In conclusion, this study provides a significant contribution to eucalypt comparative genomics, by examining differentiation between *Corymbia* species and *E. grandis*. The results reported here are the first glimpses into the changes that have occurred between two eucalypt genera since their divergence. Our experimental design and stringent methodology provides compelling evidence for chromosomal rearrangements between these genera. Despite these rearrangements our findings, together with past studies, suggest woody plants are characterised by a low rate of structural evolution in comparison to grasses and other herbaceous genera. The linkage maps constructed in this study have been crucial in the *de novo* assembly of the CCV genome (Shepherd et al. 2016), which has allowed more detailed comparative analysis of individual gene families (Butler et al. 2017a), both of which will be reported in subsequent studies.

Supplementary material

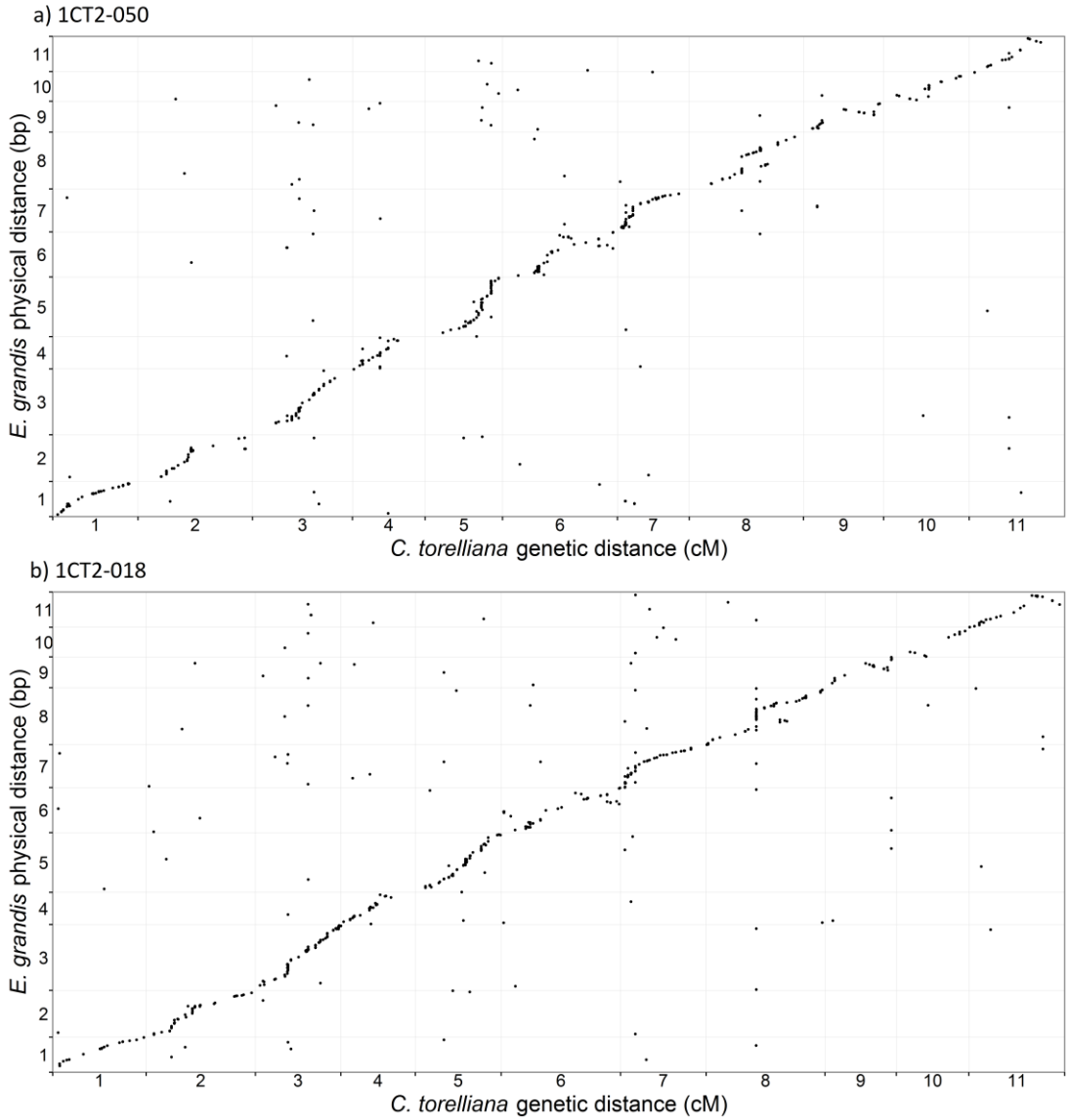
Supp. 2.1 Summary of the number of test-cross markers available for mapping in each parent of the *Corymbia* pedigrees by class and type

Marker type	Class	1CCV2-054	1CT2-050	1CT2-018
Dominant	1	1645	1934	2626
	2	5580	2449	1555
	3	760	510	556
	Subtotal	7985	4893	4737
SNP	1	126	70	93
	2	729	342	272
	3	1886	1249	1221
	Subtotal	2741	1661	1586
Total		10726	6554	6323

Supp. 2.2 Comprehensive linkage maps created for each parent of the *Corymbia* pedigrees. a) *Corymbia citriodora* subsp. *variegata* (1CCV2-054) b) *Corymbia torelliana* (1CT2-050) c) *Corymbia torelliana* (1CT2-018)

This material is unfeasible to present in this thesis due to its format. It can be found alongside the published paper (referred to as Table S2) at: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-017-3782-7

Supp. 2.3 Marker position in the *Corymbia torelliana* maps ((a) 1CT2-050 and (b) 1CT2-018) relative to the *Eucalyptus grandis* genome

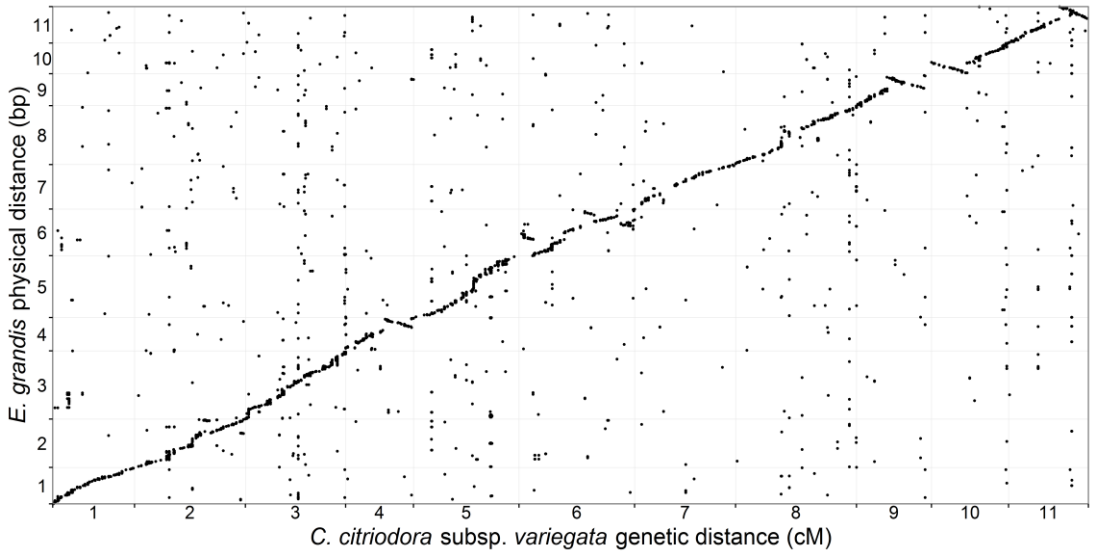


Supp. 2.4 Correlation of putatively rearranged areas of the *Eucalyptus grandis* genome with independently constructed linkage maps in *E. grandis* and *E. urophylla*.

Chromosome	*E. grandis* Spearman's correlation	Degrees of freedom	*E. urophylla* Spearman's correlation	Degrees of freedom
2	NA	NA	0.92***	23
4	0.99***	25	1.00***	11
6	0.99***	9	0.84***	9
6	0.96***	35	0.99***	15
6	0.96***	8	0.99***	10
8	0.99***	65	0.99***	45
9	0.99***	26	0.99***	15
10	0.99***	27	0.99***	30
11	0.99***	74	0.99***	93

The area spanned by the rearrangement on chromosome 2 in the *E. grandis* map only had a single marker, so orientation of that section could not be validated. The lowest correlation was on chromosome 6 in *E. urophylla* due to three marker positions that were not collinear, but these were interspersed throughout the region and suggested no gross inaccuracies in structure. *** $P < 0.001$

Supp 2.5 Duplicate marker position in the *Corymbia citriodora* subsp. *variegata* map relative to the *Eucalyptus grandis* genome.



An e-value threshold of $1e^{-10}$ gave approximately 6000 high scoring pairs, which were allowed to be matched to multiple positions. Visual inspection of dot matrixes revealed no series of collinear markers that were represented on multiple chromosomes, suggesting no instances of inter-chromosomal duplications in *E. grandis* relative to *Corymbia.*

Chapter 3 - Evidence for different QTL underlying the immune and hypersensitive responses of *Eucalyptus globulus* to the rust pathogen *Puccinia psidii*

Introduction

Puccinia psidii is a rust pathogen affecting many species in the Myrtaceae family (Coutinho et al. 1998), a predominantly southern hemisphere plant family with many economically significant species (Grattapaglia et al. 2012). *Puccinia psidii* was first described in 1884 by Winter in Brazil as guava rust (Coutinho et al. 1998). It causes lesions on actively growing young leaves and shoots which can lead to defoliation and reduced growth (Minchinton et al. 2014), as well as stem and branch dieback, and in extreme cases tree malformation and mortality (Ferreira 1983; Pegg et al. 2014b). Rust infection is optimal in moist conditions and moderate temperatures (Piza and Ribeiro 1988; Kriticos et al. 2013), and the pathogen has been reported in both temperate and tropical climates (Pegg et al. 2012). Its spread can be exacerbated by periods of high rainfall and humidity which create optimal conditions for the growth of both plants and spores and may result in a high number of new shoots and leaves becoming infected (Pegg et al. 2014b).

Puccinia psidii is a pathogen of global importance to Myrtaceae industries due to its wide host range and destructive effects (Glen et al. 2007). It has for example been detected in the USA (Coutinho et al. 1998), Hawaii (Uchida et al. 2006), Japan (Kawanishi et al. 2009), China (Zhuang and Wei 2011), South Africa (Roux et al. 2013) and Australia (Carnegie et al. 2010). Large scale losses were reported in Brazilian plantations of *Eucalyptus grandis* (Coutinho et al. 1998), where the native pathogen was able to infect the introduced species. Microsatellite analysis suggests that the strains of *P. psidii* infecting eucalypts in Brazil have been reproductively isolated from those infecting guavas for over 1,000 years, although the origin of the eucalypt-infecting strains is uncertain (Graça et al. 2013). Due to its spread and the implications for commercial plantations, rust has long been recognised as a significant economic and ecological threat to Australia's Myrtaceous dominated flora (Glen et al. 2007).

Currently, only a single strain of *P. psidii* has been detected in Australia, which was originally referred to as myrtle rust in an attempt to distinguish it from the various biotypes found in other continents (Carnegie and Lidbetter 2012). The strain present in Australia

shares the same multilocus microsatellite genotype with a strain found in the Hawaiian islands (Machado et al. 2015). This strain has not been found in eucalypt or guava in Brazil (Zhong et al. 2011), but variability is high in its native range, and it may be rare. Myrtle rust was originally detected in Australia in 2010 in a cut flower nursery on the New South Wales central coast (Carnegie et al. 2010). From the initial disease incursion, it spread rapidly and was detected in south eastern Queensland in late 2010, Victoria in late 2011 (Pegg et al. 2012), northern Tasmania in early 2015 (Tobias et al. 2015) and the Northern Territory later in 2015 (Liberato et al. unpublished). Its current distribution includes native forests along the east coast of Australia from Batemans Bay in New South Wales to Daintree, 2800 km north in Queensland, and in nurseries and private gardens around Victoria (Minchinton et al. 2014). Its current host range in Australia, already spans 347 species from 57 genera of Myrtaceae, and is very broad in comparison to other fungal pathogens (Giblin and Carnegie 2014). Australian impacts include severe damage in commercial plantations of *Backhousia citriodora* and *Syzygium anisatum*, as well as in native forests with damage to *Rhondomyrtus psidioides* and *R. rubescens* reported (Carnegie et al. 2015). Myrtle rust has also been detected in commercial plantations of *E. grandis*, but with no serious damage reported as yet (Carnegie 2015). Climate modelling has been undertaken to identify regions that are of high risk of rust invasion/infection (Kriticos et al. 2013), and these include a considerable expanse of commercial eucalypt plantations (Iglesias-Trabado et al. 2009).

The deployment of resistant genotypes is an effective way of managing such diseases in plantations. Quantitative genetic studies have revealed that significant additive genetic variation in resistance resides within commercial tree species of both *Eucalyptus* (Balmelli et al. 2014) and the sister genus *Corymbia* (Pegg et al. 2014a). However, while variation in resistance to rust has been observed between and within Myrtaceae species, little is known about the underlying molecular genetic control of this variation. In the first quantitative trait loci (QTL) study of rust in eucalypts, Junghans et al. (2003a) used RAPD linkage maps to discover the resistance QTL *Ppr1* in *E. grandis*. This QTL was later positioned on linkage group 3 using a microsatellite reference map for *Eucalyptus* (Brondani et al. 2006), and validated in additional families (Mamani et al. 2010). More recently, examination of the *E. grandis* reference genome sequence highlighted a large family of NB-LRR genes in the same genomic region as this QTL, one of which was significantly associated with variation in rust resistance in *E. grandis* (Thumma et al. 2013). Breeding programs deploying clonal genotypes began to select trees with favourable alleles at this QTL for use in plantations in areas of Brazil prone to rust (Labate et al. 2009). However, a later study found trees carrying this QTL were able to be infected with a new rust strain, suggesting that *Ppr1* was failing (Graça et al. 2011). Further study found a number of additional interacting QTL

affecting variability in rust resistance in hybrids of *E. grandis*, *E. globulus*, *E. urophylla* and *E. dunnii* (Alves et al. 2012), revealing the complexity of rust resistance. A study of gene expression in *E. grandis* has suggested that two responses may contribute to the resistant phenotype, with the genes for both cellular polarisation and systemic resistance mechanisms differentially expressed in resistant *versus* susceptible trees (Moon et al. 2007).

While native to Australia and islands to its north, eucalypts are grown around the world for diverse purposes, including pulpwood plantations (Doughty 2000) with specific species chosen for performance in different climates (Booth and Pryor 1991). *Eucalyptus globulus*, native to south eastern Australia (Dutkowski and Potts 1999), is one of the major hardwood species grown in pulpwood plantations in temperate regions of the world (Potts et al. 2004), including Australia (Gavran 2014). Accordingly it has been the subject of extensive genetic studies, including QTL studies of wood properties and growth (Freeman et al. 2009; Thumma et al. 2010; Freeman et al. 2013) and foliar chemistry (Freeman et al. 2008a). *Eucalyptus globulus* is susceptible to several diseases such as *Mycosphaerella* leaf disease caused by *Teratosphaeria cryptica* that have severe detrimental impacts on tree growth and survival and thus affect its viability as a plantation species in high disease risk areas (Pinkard et al. 2010; Balmelli et al. 2013; Hamilton et al. 2013; Kriticos et al. 2013). However, there is little information on QTL affecting disease resistance in *E. globulus* (Junghans et al. 2003a; Freeman et al. 2008b; Alves et al. 2012).

This study aims to dissect the genetic basis of variation in myrtle rust resistance in *E. globulus* through determination of the number and location of QTL underlying this variation. We used the novel technique of phenotyping open pollinated (OP) progeny collected from 218 trees of an outbred F_2 mapping family for QTL analysis. For comparison, we also mapped QTL for other fungal pathogens (Freeman et al. 2008b) as well as previously discovered QTL for myrtle rust in different species (Junghans et al. 2003a; Alves et al. 2012). Potential mechanisms that may underlie the different resistance responses observed for myrtle rust are discussed.

Material and methods

Experimental design

The QTL analysis used 218 genotypes from an outbred F_2 population, which has been previously used for linkage map construction and QTL analysis (Hudson et al. 2012a; Hudson et al. 2014). The population was generated from crossing F_1 parents, each derived from crossing unrelated trees originating from Wilsons Promontory Lighthouse (LH) with

trees originating from King Island in Bass Strait (KI) and Taranna in southeast Tasmania (TA). The following pedigree was produced: 614LH/KI440//615LH/TA423.

Phenotyping of the mapping family for rust resistance was undertaken based on the performance of open-pollinated (OP) progeny of the F_2 individuals which were reproductively mature when the trees were eight years old. The rust screening trial included 218 OP families (134 from a trial at Boyer and 84 from a Geeveston trial) as well as six control seedlots and OP seed from progeny of two of the parents of the mapping population. These were grown in a randomized incomplete block design with 15 replicates and 11 incomplete blocks per replicate. Each OP family was represented once per replicate, except in the few cases where insufficient seedlings were available and fillers from other families were used. This design was maintained throughout the experiment from seedling growth, inoculation, incubation and assessment.

Inoculum, inoculation and rust assessment

Inoculum was collected as described by Pegg et al. (2014a). In brief, a pustule isolate of *P. psidii* was collected from *Rhodamnia sessiliflora* growing in the Chapel Hill suburb of Brisbane, Queensland, Australia. Urediniospores were collected and added to 5 mL of sterile distilled water with one drop of Tween 20. This suspension was applied to *Syzygium jambos* and *R. rubescens* seedlings and covered with plastic bags to maintain high humidity and incubated at 18 °C for 24 hours. Plants were then placed in a shadehouse under natural light conditions and watered as required. Urediniospores were collected and placed in a desiccator for 5 - 7 days before being placed in Nunc tubes and stored at -80 °C. This was repeated a number of times and spores stored until required for inoculation.

Inoculation of *E. globulus* seedlings was performed as described by Pegg et al. (2014a). In brief, urediniospores were allowed to thaw and added to sterile distilled water with Tween 20 added at a rate of two drops per 100 mL of water. Spore counts were conducted with a haemocytometer, and the suspension was adjusted to a concentration of 1×10^5 spores mL^{-1}. Seedlings were inoculated using a fine mist spray applied to both the upper and lower leaf surfaces. To maintain high humidity and leaf wetness the seedlings were covered with a plastic sheet and placed in a controlled environment room at 18 - 20 °C in the dark for 24 hours. Plants were then moved to a shadehouse with disease symptom progression monitored daily.

Seedlings were assessed 20 days after inoculation for the severity of infection on new shoots and leaves using a disease rating scale modified from Junghans et al. (2003b): 1 = no symptoms evident or presence of chlorotic flecking; 2 = presence of a hypersensitive

reaction (HR) with fleck or necrosis; 3 = small pustules, < 0.8 mm diameter, with one or two uredinia; 4 = medium-sized pustules, 0.8 – 1.6 mm diameter with about 12 uredinia; 5 = large pustules, > 1.6 mm diameter, with 20 or more uredinia on leaves, petioles and/or shoots (Pegg et al. 2014a). Only seedlings which were actively growing and had new shoots and leaves at inoculation were assessed. The mean rust score for each OP family was used as a proxy for the phenotype of the F_2 parent tree.

A second round of inoculation and scoring was performed to verify the stability of the resistance scores. After the initial scoring, seedlings were cut back to approximately two leaves with no diseased tissue and allowed to regrow until two pairs of new leaves were produced with evidence of a third new flush of leaves to ensure the seedlings were not stressed. Inoculation and scoring was done as before.

Any seedlings that did not have a resistance score in both assessments, due to death or failure to produce new growth, were excluded from further analysis. Seedlings were grouped into categories of: no reaction (NR, disease score of 1); HR (disease score of 2); and presence of infection/pustules (INF, disease score 3 - 5), based on the highest state of infection recorded over the two assessments to account for incomplete infection at the time of scoring or escapes from inoculation. The mean rust score were subsequently calculated for each OP family based on the raw data. The data were also subdivided into (i) the proportion of seedlings exhibiting no reaction (i.e. NR/(NR+HR+INF)); and (ii) the proportion of seedlings with a reaction which exhibited a hypersensitive response (i.e. HR /(HR+INF)).

QTL analysis for rust

The linkage map described by Hudson et al. (2012a) was used for QTL analyses. This bi-parental consensus map contained 50 microsatellite (SSR) and 1,010 DArT markers. To reduce computational demands for QTL analyses, map reduction was performed as described in Hudson et al. (2014). In brief, this involved selecting a subset of 391 evenly spaced markers which retained full genome coverage (see Hudson et al. 2014, for more detail). The numbering and orientation of linkage groups correspond to the 11 scaffolds of the *E. grandis* reference genome sequence (v2 www.phytozome.net) (Myburg et al. 2014).

For each subdivision of data, QTL analyses were performed using the regression algorithm implemented in MapQTL 6.0 (Van Ooijen 2009) based on the proxy phenotype of 218 F_2 progeny. Permutation tests were run in MapQTL 6.0 to determine logarithm of odds (LOD) significance thresholds at genome-wide and chromosome-wide levels (1,000 permutations) (Churchill and Doerge 1994). Putative QTL were declared at two different levels, significant

(genome-wide type I error rate < 0.05) and suggestive (chromosome-wide type I error rate < 0.05). QTL analyses were first performed using interval mapping. For each putative QTL exceeding the suggestive threshold in interval mapping, the marker closest to the QTL peak were chosen as cofactors for restricted multiple-QTL model (rMQM) mapping. rMQM analyses were performed using an iterative approach until no further QTL were detected, selected cofactor markers were the closest marker to each QTL, and QTL positions were stable (Van Ooijen 2009).

To test for QTL interaction for each of the three variables, a 1:1 marker closest to each QTL peak was tested in a pair-wise manner using two-way ANOVA. Data were compared using Tukey's HSD test. These tests were undertaken using the packages 'ggplot2' (Hadley 2009) and 'multcompView' (Graves et al. 2012) of R (R Core Team 2017).

The linkage map was also searched for epistatic QTL using IciMapping (Li et al. 2008). For this analysis, the outbred F_2 cross was recoded into a pseudo-testcross and the consensus linkage map was split to create separate linkage maps for each F_1 parent tree (Van Ooijen 2009), allowing for two dimensional inclusive composite interval mapping (ICIM) analysis for epistasis. The parameters used for ICIM analyses were a 5 cM genome scanning step and a threshold LOD of 3.5, following (Alves et al. 2012). The distribution of these QTL across linkage groups was tested for departure from expectations based on the number of annotated genes across the *E. grandis* genome (Myburg et al. 2014), using a Chi-square test.

QTL analysis for other fungal diseases

The F_2 trees, as opposed to their OP progeny which were used for rust assessment, were also assessed for resistance to natural infection by other fungal diseases. This field assessment included 177 trees from Boyer and 326 from Geeveston. A small branch approximately 40 cm in length was taken from each tree and scored for severity of fungal damage as a percent of total adult leaf biomass. Leaves with spots representative of the prevalent symptoms at each of the sites were collected for DNA analysis of the fungi associated with the symptoms, which were classed according to macroscopic morphology as either *Sonderhenia*-like or *Teratosphaeria*-like. A representative spot was excised from 10 leaves for each site x spot morphology combination and placed individually into 1.5 ml microcentrifuge tubes for DNA extraction. DNA was extracted and purified according to Yuskianti et al. (2014) and fungal internal transcribed spacers (ITS) amplified by PCR and sequenced for identification. QTL analyses were undertaken as above using the phenotype

of the F_2 trees, with branch mass used as a covariate to account for heterogeneity in the size of each sample at Geeveston.

Positioning of candidate genes and other QTL

In order to identify positional candidate genes for the putative QTL in this study, we placed genes known to be involved in pathogen resistance in plants on the linkage map and compared their location to our QTL positions. These genes included those which have been implicated in pathogen resistance in plants (Thumma et al. 2013), identified by a keyword search, and those differentially expressed after exposure to rust in individuals of *E. grandis* (Moon et al. 2007), identified by gene accession search and BLAST of protein sequences against the *E. grandis* genome (v2 www.phytozome.net) (Myburg et al. 2014). MiRNA loci (miR156 and miR172 precursors) previously annotated and placed on the linkage map (Hudson et al. 2014) were also examined, since these have been implicated in disease resistance (Zhao et al. 2012). For comparative purposes, *Ppr1* (Junghans et al. 2003a) and other QTL from this (Supp. 3.1) and other pedigrees (Supp. 3.2, Freeman et al. 2008b; Alves et al. 2012) implicated in fungal resistance in eucalypts were also placed on the linkage map. The QTL from outside our pedigree were positioned based on the sequence of flanking markers.

Results

Analysis of variation in rust resistance

In the first and second inoculations, 2,784 and 2,668 seedlings were scored for resistance, respectively, with 2,469 seedlings represented in both inoculations. Of the initial 15 replicates, the mean number of seedlings per family that were scored both times was 11 (SD = 2.3). Family mean rust score ranged from 1.0 to 2.8 in both the first and second inoculation (Figure 3.1), with an overall mean rust score of 1.4 and 1.5 respectively. Resistance was relatively high in each family as the majority of individuals were scored as 1, showing no symptoms. The classification of seedlings based on the highest infection state observed in either the first or second inoculation resulted in 1,469 NR, 489 HR and 436 INF for the F_2.

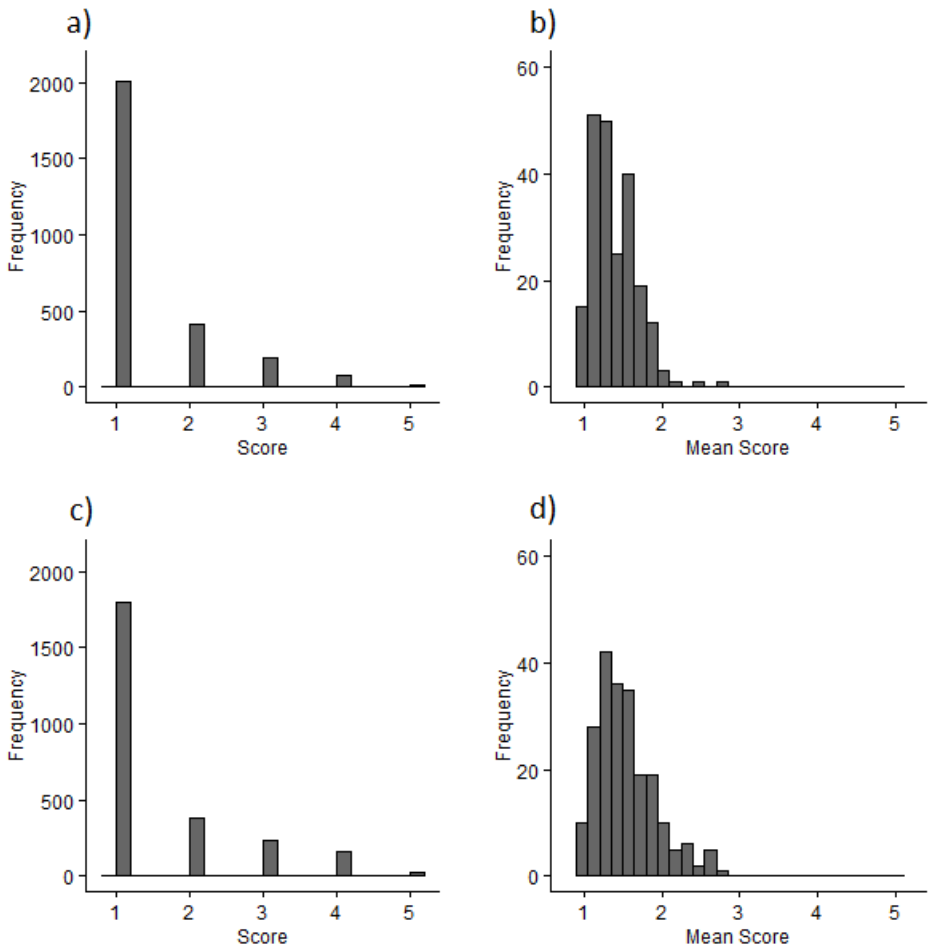


Fig. 3.1 Frequency of open-pollinated progeny of the *Eucalyptus globulus* F_2 population with various rust (*Puccinia psidii*) scores. a – Individual score frequency, first inoculation. b – Family mean score frequency, first inoculation. c – Individual score frequency, second inoculation. d – Family mean score frequency, second inoculation

There was clear evidence of genetic segregation for resistance within the F_2 family, shown by a significant difference between the rust scores in the OP families in both the first (χ^2_{217} = 339, P < 0.001) and second (χ^2_{217} = 396, P < 0.001) inoculation (based on Kruskal-Wallis ANOVA). Further, the resistance scores of the first and second inoculation were significantly positively correlated, both at the individual seedling (Pearsons r = 0.62, df = 2467, P < 0.001) and family mean (Pearsons r = 0.74, df = 224, P < 0.001) levels. Most seedlings remained in the same category (NR, HR or INF) over the two inoculations (1809 or 73%), while 660 had unstable resistance scores and shifted category. These shifts were generally minor, with 537 individuals only shifting a single category in our scale. The control families selected for their extreme susceptibility and resistance displayed the relevant mean phenotype, indicating the repeatability of the resistance response.

QTL for disease resistance

Two putative QTL were identified for the proportion of NR seedlings, one at the suggestive level and one at the significant level (Table 3.1, Figure 3.2). These QTL were located on linkage groups 3 (*Ppr2*) and 7 (*Ppr3)*. Together, these QTL explained an estimated 32% of the phenotypic variance in this trait. Two further putative QTL were identified for the proportion of HR seedlings (i.e. excluding NR trees), both at the significant level (Table 3.1, Figure 3.2). These two QTL were located on linkage groups 6 (*Ppr4*) and 9 (*Ppr5*) and together explained an estimated 25.1% of the phenotypic variance in this trait. QTL analyses undertaken with the raw data, i.e. the family mean over both assessments without subdividing the data, detected two of the same QTL (*Ppr2* and *Ppr4*), but at a lower LOD. *Ppr3* and *Ppr5* failed to reach significance in this analysis, suggesting additional power had been gained by splitting the resistance response into the NR and HR classes.

Table 3.1 Putative QTL for rust (*Puccinia psidii*) resistance identified by MQM mapping in the *Eucalyptus globulus* F_2 population

QTL	Category	LG	cM[a]	Adj. marker[b]	Flanking markers[c]	LOD[d]	PVE[e]	Physical position (bp)[f]		SEG[g]
								Peak	Flanking	
Ppr2	NR	3	50.1	572474	568396	15.79***	27.2	37,394,764	28,711,882	F
					572802				42,683,460	
Ppr3	NR	7	50.8	570240	504063	3.09	4.6	45,281,885	31,587,467	M
					566156				45,879,882	
Ppr4	HR	6	23.4	504554	641098	10.46***	18.2	14,701,365	9,239,564	F
					639232				17,425,979	
Ppr5	HR	9	17.3	566438	640922	4.24*	6.9	11,150,046	*1,718,324*[b]	F
					566701				23,366,994	

NR no reaction, HR hypersensitive reaction, LG linkage group

[a] QTL LOD peak position

[b] Adjacent marker to QTL LOD peak. In the case of the first flanking marker of *Ppr5* the sequence of the closest marker was located on a minor scaffold, so the physical position of the closest adjacent marker anchored to the main scaffolds is reported.

[c] Markers flanking outside of 2 LOD confidence interval from QTL peak

[d] Peak LOD score for each QTL. Genome-wide significance is indicated by *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$. The remaining QTL were significant at the suggestive level (chromosome-wide type I error rate < 0.05).

[e] The proportion of phenotypic variation explained by each QTL

[f] Midpoint of marker sequence based on BLAST against *Eucalyptus grandis* genome v2

[g] Segregation of the QTL effect (M=male, F=female)

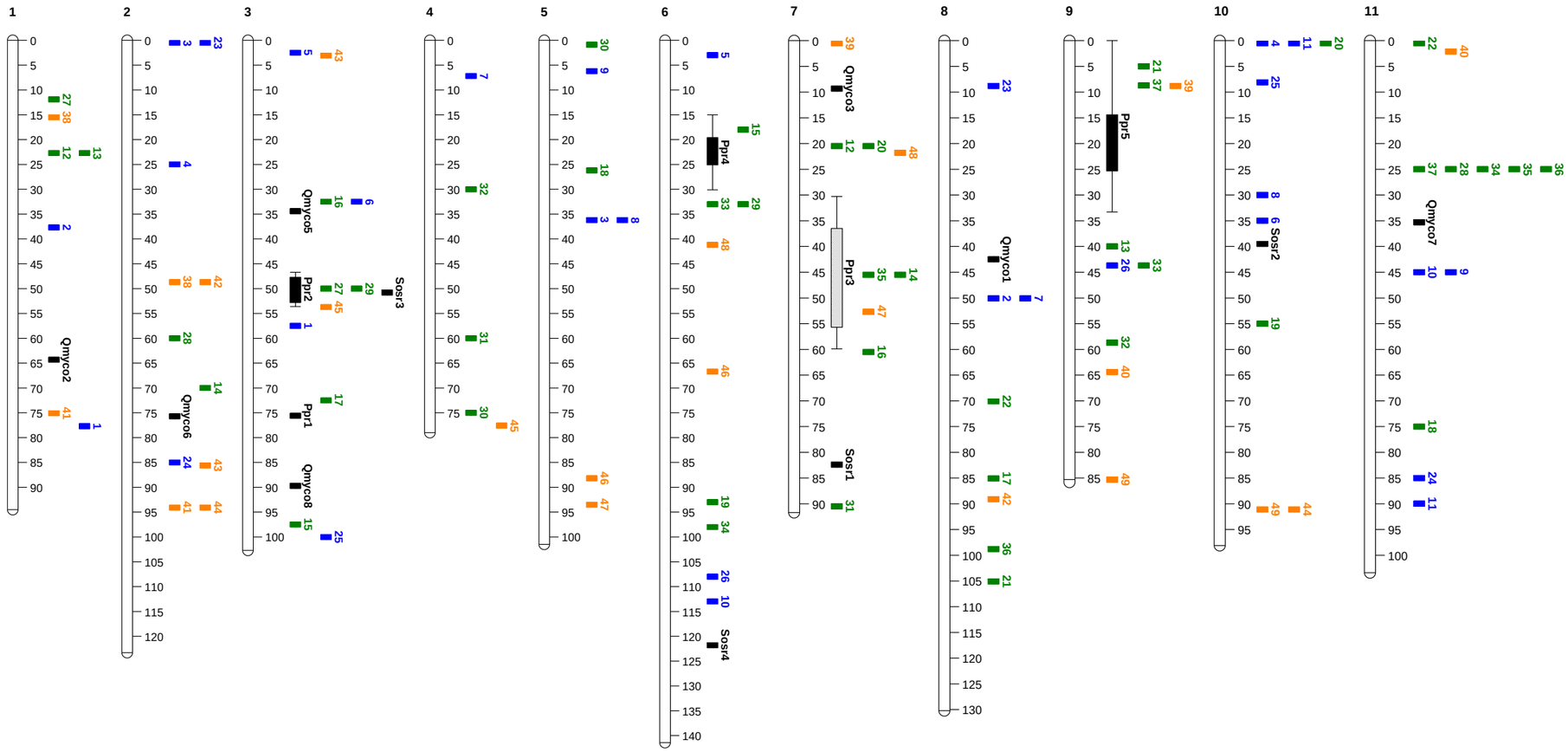


Fig. 3.2 The location of fungal resistance QTL on the *Eucalyptus globulus* linkage map. Bars and lines indicate one-LOD and two-LOD QTL confidence intervals respectively. Hashed bar indicates QTL at the suggestive level. Previously published QTL for resistance to fungal pathogens (*Ppr1* (Junghans et al. 2003a; Alves et al. 2012), *Qmyco1 - 5* (Supp. 3.2, Freeman et al. 2008b)) plus unpublished QTL (*Qmyco6 - 8* and *Sosr1 - 4,* Supp. 3.1) shown as QTL peak position. Digenic interacting QTL are indicated by matching number and shown as QTL peak position. Blue numbers indicate HR QTL, green indicate NR QTL, and orange indicate previously published interacting QTL for *Eucalyptus* hybrids (Alves et al. 2012). 1 - 22 discovered through analysis of male map, 23 - 37 through female map, and 38 - 49 published in Alves et al. (2012).

Of the six tests for interaction between our four QTL, two were significant. The interaction between the two QTL *Ppr2* and *Ppr4* was highly significant in the NR data ($F_{1216} = 14.57$, $P < 0.001$), suggesting epistasis between these QTL. Analysis of the genotype means at these QTL suggests that the effect of the susceptible allele at *Ppr4* is completely masked by the resistant allele at *Ppr2*, implying that *Ppr2* takes precedence in determining a NR phenotype (Figure 3.3). The interaction between the two QTL *Ppr3* and *Ppr5* was also significant ($F_{1216} = 6.5$, $P < 0.05$), but no masking effect was observed.

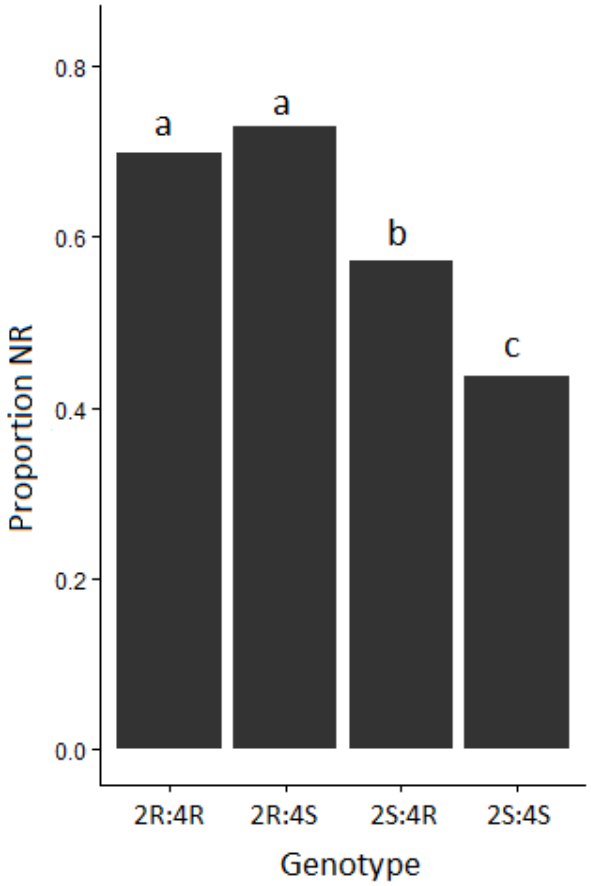


Fig. 3.3 Genotype means for different combinations of alleles at *Ppr2* and *Ppr4* (*Ppr2:Ppr4*), for the NR phenotype. Proportion NR refers to the mean proportion of seedlings within each family showing no response (immune response). The presence of allele 4S (susceptible) at *Ppr4* is masked by allele 2R (resistant) at *Ppr2*

Seven putative QTL for other pathogens were also identified using phenotype information directly obtained from the F_2 trees growing in the field trials at Boyer and Geeveston (Figure 3.2). Three QTL for resistance to a pathogen causing *Teratosphaeria*-like symptoms (*Qmyco6 - 8*) were located on linkage groups 2, 11 and 3 respectively, and four QTL for resistance to a pathogen causing *Sonderhenia*-like symptoms (*Sosr1 - 4*) were located on linkage groups 7, 10, 3 and 6 respectively (Supp. 1.1 for full details).

Epistatic QTL

Using two dimensional ICIM mapping for the proportion of plants with no reaction (i.e. NR/NR+HR+INF), we detected 13 pairs of interacting QTL based on the paternal map and 12 pairs on the maternal map (Figure 3.2, Supp. 3.5). For the proportion of plants with hypersensitive response (i.e. HR/HR+INF) we detected 13 and five QTL pairs based on the paternal and maternal map, respectively. The phenotypic variance explained by the various pair-wise epistatic QTL ranged from 1.9 - 31.6%, with some interacting with the major additive effect QTL *Ppr2* and *Ppr4*. The interaction reported earlier between *Ppr2* and *Ppr4* was also detected here, while the less significant interaction between *Ppr3* and *Ppr5* was not. After removing QTL within ± 1.25 cM of each other within each trait (as these potentially represent the same QTL), 57 unique epistatic QTL were found, three of which (5.3%) affected both traits. The number of QTL discovered on each linkage group in this study did not significantly depart from expectations based on the number of annotated genes across linkage groups on the *E. grandis* genome ($\chi^2_{10} = 11.4$, $P > 0.1$) (Myburg et al. 2014), suggesting no bias towards resistance loci to be on any linkage group.

Candidate genes/co-location

Physical positioning of genes that may be implicated in resistance (Thumma et al. 2010) identified 153 positional candidate genes located within the confidence intervals of our four QTL. Twenty nine genes were located within the confidence interval of *Ppr2* (with 9 within ± 1 MB of the QTL peak), 60 within *Ppr4* (22 ± 1 MB of peak), 30 within *Ppr3* (12 ± 1 MB of peak) and 34 within *Ppr5* (2 ± 1 MB of peak) (Supp. 3.3). While some other potential candidates were evident, each QTL confidence interval was dominated by a cluster of genes annotated with similar functions: *Ppr2, Ppr4* and *Ppr5* each contained gene clusters from the large gene families annotated as NB-LRR and NB-ARC, and *Ppr3* contained a cluster of cyclase/dehydrase and lipid transport genes. Other genes found under the confidence intervals include MYB and MLP-like proteins, glycosyl hydrolase proteins, bZIP transcriptional factors and peroxidase proteins (Supp. 3.3).

Amongst the genes that were differentially expressed in individuals of *E. grandis* that were resistant *versus* susceptible to rust (Moon et al. 2007), six fell within our QTL confidence intervals. Four of these genes were over-expressed in susceptible trees, and of these one gene fell within the confidence interval of *Ppr2*, two within *Ppr3*, and one within *Ppr5*. Two genes over-expressed in resistant trees also fell within *Ppr5* (Supp. 3.4).

Of the QTL for other pathogens discovered in this pedigree, *Sosr3* was co-located with *Ppr2* (within 0.7 cM of the QTL peak). Repositioning of previously published QTL for

Teratosphaeria resistance in *E. globulus* (Supp. 3.2, Freeman et al. 2008b) found no co-location with our rust QTL. Of the 24 epistatic QTL for rust resistance identified by Alves et al. (2012) in *Eucalyptus* interspecific hybrids, QTLs "43" and "39" (as labelled in our analysis) co-located with the QTLs "24" and "37" detected in this study (i.e. within ± 1.25 cM), but no interacting pairs were co-located.

Discussion

Traditionally in forestry, phenotyping for QTL studies is performed directly on the trees being used in the analyses (Freeman et al. 2008b; Hudson et al. 2014), or on clonal replicates (Zobel 1993; Hüberli et al. 2001). In this study we chose instead to employ progeny testing, which is commonly used in breeding animals and plants (Kashi et al. 1990; Wenzel and Foroughi-Wehr 1990), including forest trees (Williams et al. 1987; Johnson and Burdon 1990), and has been used previously in QTL studies employing controlled breeding in cattle (Georges et al. 1995). To our knowledge, this is the first study to use progeny testing for QTL mapping purposes based on OP seed collected from a controlled cross. Our results demonstrate that progeny testing is an effective method of phenotyping established trees in mapping populations. This technique offers several distinct advantages over conventional phenotyping in QTL studies in trees, which either have no replication (single tree), or require expensive cloning to produce replication (Bradshaw and Stettler 1995; Wu 1998). Progeny testing of mapping families extends the use of the maps through phenotyping traits that are otherwise impractical to assay in established trees, such as juvenile expressed traits, traits that are impossible to measure due to quarantine or other restrictions, and those that are better measured in controlled glasshouse conditions such as physiological traits. Progeny testing of mapping families will also prove useful in studying genotype by environment interactions (e.g. Freeman et al. 2013) since the OP families can be planted in multiple environments and across environmental gradients.

This is the first report of QTL influencing susceptibility to *P. psidii* in pure *E. globulus*, building on past reports in other eucalypt species (Junghans et al. 2003a; Mamani et al. 2010; Rosado et al. 2010; Alves et al. 2012). Four QTL were described underlying rust resistance in this *E. globulus* family, which did not overlap with the previously reported locus *Ppr1* in *E. grandis* (Junghans et al. 2003a). Together with the discovery of numerous interacting QTL with no additive effects in this pedigree, our findings support the hypothesis that rust resistance in eucalypts is complex and influenced by many loci, in line with findings in other eucalypts, including *E. grandis* (Mamani et al. 2010; Thumma et al. 2013), *E. pellita* (Santos et al. 2014) and interspecific hybrids (Alves et al. 2012).

The different resistance responses exhibited in this pedigree (immune or no response (NR) and hypersensitive response (HR)) may reflect different modes of defence against pathogens. Plant defences against pathogens include preformed physical and chemical barriers to infection, and induced defences based on pathogen recognition (Xiao et al. 2008). While exceptions have been noted (Thomma et al. 2011), induced responses are commonly classified into two main mechanisms: pathogen associated molecular pattern triggered immunity (PTI), and effector-triggered immunity (ETI) (Jones and Dangl 2006; Dodds and Rathjen 2010). PTI operates via pattern recognition receptors, recognising common components of all pathogens (such as fungal chitin and bacterial flagellin) and inducing a defence response. Pathogens are able to enter the cell, release virulence molecules to suppress the PTI response, and the presence of these molecules can be recognised by the plant and induce an ETI response, which generally (but not always) triggers hypersensitivity (Jones and Dangl 2006; Cui et al. 2015).

Given the above, it is possible that the different QTL underlying the immune and hypersensitive responses in this pedigree may reflect loci involved in different resistance mechanisms. For example, the QTL influencing the NR phenotype may reflect basal defenses such as loci involved in the PTI response, which is generally symptomless (Klement et al. 2003; Jones and Dangl 2006). It may also reflect a modulated ETI response that has not reached the level of hypersensitivity (discussed below). Alternatively, the QTL underlying the NR phenotype may influence constitutive morphological or chemical traits rather than induced pathogen resistance, as no specific or generalized plant response would be expected when the pathogen is unable to penetrate the cuticle and infect cells. For example, the waxy cuticle of eucalypts has been shown to inhibit fungal growth (Martin and Juniper 1970; Jenkins and Suberkropp 1995; Canhoto and Graça 1999), as have the terpenes present in leaves (Batish et al. 2008; Liu et al. 2008).

The QTL underlying HR may reflect loci involved in an ETI response based on pathogen recognition. This hypothesis has some caveats; under the original gene-for-gene model of plant resistance which assumes co-evolution between a resistance (R) and avirulence (AVR) gene (Flor 1942), it is unlikely the eucalypts responding to the rust with HR are exhibiting ETI, as *P. psidii* is an exotic pathogen and the naïve host has had no chance for AVR/R gene co-evolution. However, while such simple pair-wise interactions between resistance and avirulence genes are still supported in some systems (Dodds et al. 2006), other interactions have been shown to be more complex (Keith and Mitchell-Olds 2013; Cui et al. 2015; Lee and Yeom 2015). For example, the guard model (Dangl and Jones 2001) could account for HR triggered by ETI, as under this model *R* genes detect modifications by effectors rather

than the effectors themselves, as seen in *Arabidopsis* (DeYoung et al. 2012) and *Nicotiana* sp. (Lu et al. 2015); which does not require co-evolution to have taken place. Alternatively, the HR response may instead be triggered by PTI, for which there is precedence in naïve hosts (Heath 2000a).

A significant interaction was detected between the QTL on LG3 (NR, *Ppr2*) and the QTL on LG6 (HR, *Ppr4*, Figure 3.3), whereby the effect of the susceptible allele at *Ppr4* was masked by the resistance allele at *Ppr2*. This finding suggests the mechanism controlling the NR phenotype acts prior to the mechanism controlling the HR phenotype. If the NR mechanism is a morphological or chemical trait as discussed above, the resistant allele at that locus may prevent the pathogen from entering the plant cells, allowing no specific resistance response to develop. Alternatively, if the NR mechanism is part of an induced response, it would suggest that there is a progression of intensity of response to the pathogen (such as the 'zig-zag' model proposed by Jones and Dangl (2006)) and the mechanism underlying NR is earlier in the pathway compared to HR (Crabill et al. 2010). Induced resistance responses inhibiting the HR have been observed in both *Capsicum annuum* infected with *Xanthomonas campestris* (Newman et al. 2000), and *Nicotinia tabacum* infected with *Pseudomonas syringae* (Klement et al. 2003).

Several digenic interacting QTL for rust resistance have been previously reported (Alves et al. 2012), however there were very few that co-located with the epistatic QTL reported in this study. This result is not unexpected, as QTL studies will only identify QTL that are segregating in the cross examined, and co-location of interacting pairs between studies would require both members to be polymorphic in the genetic material used in each study and for their effects to be conserved in different genetic backgrounds. The lack of co-location also may reflect different loci underlying resistance in *E. globulus* and *E. grandis,* as an initial association genetic study suggests there is little overlap in loci affecting rust resistance between *E. globulus* and *E. grandis* when tested against the same strain of the pathogen (B. Thumma, pers. comm.). A further confounding factor which may contribute to the detection of different QTL is the different strains of the pathogen used between studies (Alves et al. 2012), particularly if the mechanisms underlying host resistance involves some form of ETI, which can be strain specific. Clearly, further research is required to elucidate the degree to which the molecular basis of host resistance is conserved between eucalypt species.

A potential case of common loci influencing rust resistance across species (and pathogens) occurs on LG3 in the vicinity of our most significant QTL, *Ppr2*. Within the relatively narrow confidence interval of this QTL (Figure 3.2), a gene overexpressed in *E. grandis* susceptible

to rust was found (Moon et al. 2007), as well as a QTL involved in digenic epistasis for rust resistance in *E. dunnii* x *E. grandis* hybrids (Alves et al. 2012). *Sosr3*, a QTL for another pathogen (Supp. 3.1), also mapped very close to *Ppr2* (the peak of each QTL was separated by 0.7 cM) in our study. Such clustering of QTL could potentially reflect the pleiotropic effects of a locus with generalised effects on disease resistance, or the action of linked genes (see below) (Wang et al. 2001; Chu et al. 2004). In the case of *Sosr3* and *Ppr2*, the QTL segregated from different parents, arguing these QTL are likely to reflect the action of discrete loci (Table 3.1, Supp. 3.1). Further study will be required to differentiate between pleiotropy and linkage for the remaining QTL located in this region, although examination of the candidate genes in this region highlights some potential explanations.

Physical positioning of genes that may be implicated in disease resistance on our linkage map provided some promising candidates within the confidence intervals of our four main QTL. Specifically, the NB-LRR gene family were the major class of gene linked to disease resistance (under the functional categories identified in Thumma et al. 2013) within the confidence intervals of *Ppr2, Ppr4* and *Ppr5*. This finding is consistent with the results of a recent association genetic study, in which half of the loci significantly associated with resistance to rust in *E. grandis* were NB-LRR loci (Thumma et al. 2013). Classic *R* genes generally code for NB-LRR proteins which respond, either directly or indirectly, to the virulence effectors released by pathogens (Ellis and Jones 1998; McDowell and Woffenden 2003; Keith and Mitchell-Olds 2013). Such genes are commonly associated with the hypersensitive response (Caplan et al. 2008), but are also frequently shown to confer resistance without triggering cell death (Cui et al. 2015; Lee and Yeom 2015) making them plausible candidates for the NR (*Ppr2*) and HR (*Ppr4* and *Ppr5*) responses. In the case of *Ppr2*, the large family of NB-LRR loci could account for the 'QTL cluster' in this genomic region on LG3, with different loci potentially underlying the QTL for different pathogens and across host species

Alternatively, a microRNA156 (miR156) precursor (*EgrMIR156.4*) located 0.1 cM from the peak of *Ppr2* (Hudson et al. 2014) could potentially contribute to the QTL effects at this loci and, more broadly, the generalised effects of the QTL cluster in this region. MiRNAs regulate genes through post transcriptional silencing of mRNA, which often affects broad gene networks (Hobert 2008). The role of MiR156 as a 'master regulator' of vegetative phase change in plants is now well established (Poethig 2009; Wu et al. 2009), and evidence is accumulating that MiR156 also plays an important role in general stress responses, including to abiotic stressors (Zhang et al. 2006; Zhao et al. 2012; Cui et al. 2014). Specifically, miR156 indirectly regulates induced resistance responses by regulating the

action of NB-LRR loci through interaction with other genes and molecules (Zhai et al. 2011; Padmanabhan et al. 2013; Baldrich et al. 2015). Interestingly, precursors of miR156 and the related miR172 were also found within 0.5 cM of the peaks of *Ppr1* and *Sosr2*, consistent with a generalized role in induced resistance responses.

A number of transcription factors are also positional candidates for the rust resistance responses. Transcription factors belonging to the WRKY, bZIP and MYB families were found within the confidence intervals of our QTL, in line with evidence that loci from these families are associated with variation in rust resistance in *E. grandis* (Thumma et al. 2013). These families are involved in the expression and regulation of various genes within the defence pathways (Eulgem and Somssich 2007; Alves et al. 2013; Tsuda and Somssich 2015) making them plausible candidates for both of the resistance responses we observed.

Exotic pathogens can have devastating effects on populations of naïve hosts, as previous incidents of anthropogenic introduction have seen pathogens cause severe population decline and threat of extinction in many tree species, which in turn can result in economic impacts (for a review, see Anderson et al. 2004). This study has shown there is genetic variation for resistance to rust in this family of *E. globulus*. Together with the high levels of resistance found in other provenances of *E. globulus* from Victoria, Flinders Island and Tasmania (Zauza et al. 2010), these findings are promising for the species in respect to both conservation and the plantation industry. However, there is a lot of variability within and between species, in both *Eucalyptus* (Morin et al. 2012) and other Myrtaceous flora (Pegg et al. 2014a), suggesting that species and provenances will need to be assessed on a case-by-case basis for their response to rust (Zauza et al. 2010).

When breeding *E. globulus* for resistance to rust, selecting for multiple resistance mechanisms such as evidenced here would likely be most effective. Testing against additional strains of the pathogen, particularly the strain that is widespread on eucalypts in Brazil, would also provide additional information about the likely durability of the resistance genes. Such information is worth considering, since mutations are occurring in the strain presently in Australia (Machado et al. 2015), while the possibility of an incursion of other strains remains. In terms of potential practical applications, the mechanism underlying NR and its masking effect on HR could create difficulties in selecting for both responses using traditional phenotype based breeding techniques. Under these circumstances, molecular breeding techniques, such as marker assisted selection would be effective and should be considered.

In conclusion, we have reported QTL and identified positional candidate genes, potentially underlying multiple mechanisms for resistance to *P. psidii* in *E. globulus*. Our findings add to previous studies highlighting the complex nature of the genetic control of resistance. In view of this complexity, and the apparent species specificity of many of the loci conferring host resistance, further study to characterise the genetic architecture underlying resistance in *E. globulus* and other eucalypts would be of value.

Supplementary material

Supp. 3.1 Putative QTL for fungal diseases identified on adult foliage in the *Eucalyptus globulus* F_2 population

Trait	QTL	Population[a]	LG[b]	cM[c]	Adj. marker[d]	Flanking markers[e]	LOD[f]	PVE[g]	Physical position (bp)[h]		SEG[i]
									Peak	Flanking	
Unknown (*Sonderhenia* spp. symptoms)	*Sosr1*	G	7	82.4	564376	571617 EMBRA98	5.88***	8.0	14,867,281	561,284 31,616,752	B
	Sosr2	G	10	39.5	Embra153	EMBRA385 567691	3.16	4.2	13,717,746	9,072,464 24,188,567	B
	Sosr3	B	3	50.8	504787	639640 573637	4.63*	12.0	30,174,953	26,946,071 48,721,500	M
	Sosr4	B	6	121.8	570163	575359 565994	3.47	8.9	48,487,073	46,037,646 49,277,146	F
Teratosphaeria cryptica	*Qmyco6*	G	2	79.5	571804	503579 569164	3.83	5.3	45,847,432	41,086,661 51,343,034	F
	Qmyco7	G	11	36.6	Eg99	637129 563340	3.16	4.3	17,349,501	13,233,133 22,245,400	B
	Qmyco8	B	3	89.7	575308	570139 640146	3.08	8.8	75,010,188	63,433,248 83,090,127	F

[a] G = Geeveston, B = Boyer

[b] Linkage group

[c] QTL LOD peak position

[d] Closest marker to QTL LOD peak with sequence information

[e] First markers with sequence information flanking outside of QTL 2 LOD confidence interval

[f] Peak LOD score for each QTL. Genome-wide significance is indicated by *** P < 0.001, ** P < 0.01, * P < 0.05. The remaining QTL were significant at the suggestive level (chromosome-wide type I error rate < 0.05).

[g] The percentage of phenotypic variation explained by each QTL

[h] Midpoint of marker sequence based on BLAST against *Eucalyptus grandis* genome v2

[i] Segregation of the QTL effect (M = male, F = female, B = both)

Supp. 3.1 (cont.)

Identification of the fungi associated with the leaf symptoms was carried out as per the methods section. Sequencing of fungal internal transcribed spacers (ITS) produced several samples with a noisy chromatogram indicating the presence of several fungi, none of which dominated. A BLAST search of the International Nucleotide Sequence Databases was conducted with the trimmed and edited DNA sequences. Sequences with > 99% similarity to *Teratosphaeria cryptica* were obtained from the majority of leaves showing *Teratosphaeria*-like symptoms, collected from Geeveston. However, this species was not detected in leaves with similar symptoms collected from Boyer, which were dominated instead by a *Hormonema* sp. with > 99% sequence similarity to an endophyte isolated from leaves of *Eucalyptus globulus* in Spain (Márquez et al. 2011).

Leaves with symptoms scored as *Sonderhenia*-like were not associated with any one fungus. Most of the leaves from Boyer produced noisy chromatograms but the *Hormonema* sp. was detected in one sample. The leaves from Geeveston were colonised by a range of fungi with <95% sequence similarity to identified species. The most common was detected in two of the 10 leaves and had up to 90% sequence similarity to a range of Botryosphaeriale. Other Fungi detected included one with 95% similarity to *Celosporium larixicola,* an endophyte of larch, found in two leaves. Another had 97% similarity to *Chaetomidium arxii,* an endophyte of milk-thistle. A third species was also detected in two leaves and had up to 90% sequence similarity to a range of Botryosphaeriales. The lack of a consistent association between fungal species and symptoms, and the high similarity to endophytic species indicates that the fungi detected here are most likely to be endophytes rather than pathogens and that the leaf damage was not caused by a fungus, but possibly an insect.

Supp. 3.2 Putative QTL for *Teratosphaeria cryptica* (previously known as *Mycosphaerella cryptica*) resistance identified by MQM mapping in the 'Woolnorth' *Eucalyptus globulus* F_2 population.

QTL	LG[a]	cM[b]	Adj. marker[c]	LOD[d]	PVE[e]	Physical position (bp)[f]	SEG[g]
Qmyco1	8	42.7	567895	21.2***	34.2	17,839,133	F
Qmyco2	1	64.4	565878	7.2***	8.4	28,086,282	F
Qmyco3	7	23.3	566500	8.5***	10.2	16,371,891	F
Qmyco5	3	40.7	644467	3.2	3.5	17,563,929	M

[a] Linkage group
[b] QTL LOD peak position
[c] Closest marker to QTL LOD peak with sequence information
[d] Peak LOD score for each QTL. Genome-wide significance is indicated by *** P < 0.001, ** P < 0.01, * P < 0.05. The remaining QTL were significant at the suggestive level (chromosome-wide type I error rate < 0.05).
[e] The percentage of phenotypic variation explained by each QTL
[f] Midpoint of marker sequence based on BLAST against *Eucalyptus grandis* genome v2
[g] Segregation of the QTL effect (M = male, F = female)

Qmyco1 - 5 were originally published in Freeman et al. (2008). The results presented here are based on re-analysis of their data, using a revised linkage map (as reported in Gosney et al. 2015) with greater genome coverage and many more markers with known position on the *E. grandis* genome sequence. QTL analysis was performed as reported in this study, but was based on field infection disease severity scores from 112 clonally replicated (two individuals per genotype) genotypes, as reported in (Freeman et al. 2008). The QTL '*Qmyco4*' reported by Freeman et al. (2008) was not significant in this analysis.

Supp. 3.3 Candidate genes within rust QTL confidence intervals based on keyword search of gene classes implicated in resistance from Thumma et al. (2013)

LG[a]	Start (bp)	End (bp)	Gene name[b]	Annotation[c]	±1MB from QTL peak
3	29,516,369	29,521,138	Eucgr.C01740	ascorbate peroxidase 6	
	30,858,702	30,859,842	Eucgr.C01669	Glycosyl hydrolase family protein with chitinase insertion domain	
	30,884,647	30,886,060	Eucgr.C01666	Glycosyl hydrolase family protein with chitinase insertion domain	
	31,155,681	31,164,242	Eucgr.C01654	transmembrane receptors;ATP binding	
	32,908,472	32,909,569	Eucgr.C01980	Glycosyl hydrolase family protein with chitinase insertion domain	
	32,934,034	32,935,757	Eucgr.C01979	Glycosyl hydrolase family protein with chitinase insertion domain	
	32,947,651	32,948,748	Eucgr.C01978	Glycosyl hydrolase family protein with chitinase insertion domain	
	32,956,709	32,957,890	Eucgr.C01977	Glycosyl hydrolase family protein with chitinase insertion domain	
	33,172,237	33,178,313	Eucgr.C01968	Disease resistance protein (TIR-NBS-LRR class) family	
	33,788,800	33,789,808	Eucgr.C01943	WRKY DNA-binding protein 40	
	33,984,045	33,988,186	Eucgr.C01933	Disease resistance protein (TIR-NBS-LRR class) family	
	34,099,725	34,102,464	Eucgr.C01928	Disease resistance protein (TIR-NBS-LRR class) family	
	34,287,403	34,289,980	Eucgr.C01921	disease resistance protein (TIR-NBS-LRR class), putative	
	34,304,842	34,306,152	Eucgr.C01920	Disease resistance protein (TIR-NBS-LRR class) family	
	34,421,305	34,424,147	Eucgr.C01915	Disease resistance protein (TIR-NBS-LRR class) family	
	34,441,037	34,443,770	Eucgr.C01912	Disease resistance protein (TIR-NBS-LRR class) family	
	37,063,885	37,064,596	Eucgr.C02017	CAP superfamily protein	*
	37,536,893	37,537,872	Eucgr.C02038	Disease resistance protein (TIR-NBS-LRR class) family	*
	37,543,174	37,548,246	Eucgr.C02040	transmembrane receptors;ATP binding	*
	37,557,130	37,564,990	Eucgr.C02041	transmembrane receptors;ATP binding	*
	38,177,064	38,182,457	Eucgr.C02062	Disease resistance protein (TIR-NBS-LRR class) family	*
	38,292,983	38,295,479	Eucgr.C02064	Disease resistance protein (TIR-NBS-LRR class) family	*
	38,309,526	38,312,466	Eucgr.C02067	disease resistance protein (TIR-NBS-LRR class), putative	*
	38,515,903	38,521,663	Eucgr.C02074	disease resistance protein (TIR-NBS-LRR class), putative	*

	38,600,042	38,602,658	Eucgr.C02081	Disease resistance protein (TIR-NBS-LRR class) family	*
	38,974,914	38,977,495	Eucgr.C02096	myb domain protein 3	
	39,132,898	39,135,328	Eucgr.C02108	Disease resistance protein (TIR-NBS-LRR class) family	
	39,175,092	39,177,700	Eucgr.C02111	Disease resistance protein (TIR-NBS-LRR class) family	
	39,197,162	39,201,553	Eucgr.C02112	transmembrane receptors;ATP binding	
6	9,102,983	9,105,502	Eucgr.F00616	NB-ARC domain-containing disease resistance protein	
	9,156,274	9,162,386	Eucgr.F00621	WRKY family transcription factor family protein	
	9,187,749	9,192,786	Eucgr.F00625	tornado 1	
	9,595,463	9,596,493	Eucgr.F00659	ethylene responsive element binding factor 3	
	10,610,307	10,615,073	Eucgr.F00740	WRKY family transcription factor	
	11,049,080	11,051,833	Eucgr.F00782	NB-ARC domain-containing disease resistance protein	
	11,072,013	11,074,326	Eucgr.F00786	NB-ARC domain-containing disease resistance protein	
	11,233,148	11,234,800	Eucgr.F00791	NB-ARC domain-containing disease resistance protein	
	12,031,439	12,035,341	Eucgr.F00823	NB-ARC domain-containing disease resistance protein	
	12,063,243	12,065,927	Eucgr.F00825	LRR and NB-ARC domains-containing disease resistance protein	
	12,088,570	12,091,131	Eucgr.F00828	NB-ARC domain-containing disease resistance protein	
	12,103,021	12,109,035	Eucgr.F00830	NB-ARC domain-containing disease resistance protein	
	12,144,408	12,147,016	Eucgr.F00834	NB-ARC domain-containing disease resistance protein	
	12,177,971	12,182,949	Eucgr.F00837	NB-ARC domain-containing disease resistance protein	
	12,256,842	12,258,767	Eucgr.F00839	NB-ARC domain-containing disease resistance protein	
	12,536,267	12,539,405	Eucgr.F00872	S-locus lectin protein kinase family protein	
	12,715,501	12,720,373	Eucgr.F00886	NB-ARC domain-containing disease resistance protein	
	12,786,443	12,789,622	Eucgr.F00892	NB-ARC domain-containing disease resistance protein	
	12,831,642	12,835,627	Eucgr.F00896	NB-ARC domain-containing disease resistance protein	
	12,872,223	12,874,852	Eucgr.F00899	NB-ARC domain-containing disease resistance protein	
	12,977,478	12,981,450	Eucgr.F00904	LRR and NB-ARC domains-containing disease resistance protein	
	13,074,661	13,075,260	Eucgr.F00908	LRR and NB-ARC domains-containing disease resistance protein	

13,152,768	13,155,705	Eucgr.F00914	NB-ARC domain-containing disease resistance protein	
13,175,266	13,178,514	Eucgr.F00916	NB-ARC domain-containing disease resistance protein	
13,190,004	13,190,768	Eucgr.F00918	NB-ARC domain-containing disease resistance protein	
13,302,000	13,303,319	Eucgr.F00923	Disease resistance protein (CC-NBS-LRR class) family	
13,393,713	13,397,141	Eucgr.F00927	LRR and NB-ARC domains-containing disease resistance protein	
13,462,423	13,463,676	Eucgr.F00935	Disease resistance protein (CC-NBS-LRR class) family	
13,485,795	13,487,360	Eucgr.F00937	NB-ARC domain-containing disease resistance protein	
13,498,579	13,502,541	Eucgr.F00938	LRR and NB-ARC domains-containing disease resistance protein	
13,672,291	13,675,546	Eucgr.F00948	LRR and NB-ARC domains-containing disease resistance protein	
13,750,115	13,753,302	Eucgr.F00954	NB-ARC domain-containing disease resistance protein	*
13,775,617	13,784,018	Eucgr.F00956	NB-ARC domain-containing disease resistance protein	*
14,242,431	14,245,363	Eucgr.F01009	NB-ARC domain-containing disease resistance protein	*
14,286,428	14,289,427	Eucgr.F01014	NB-ARC domain-containing disease resistance protein	*
14,318,765	14,322,710	Eucgr.F01017	NB-ARC domain-containing disease resistance protein	*
14,338,682	14,341,375	Eucgr.F01019	NB-ARC domain-containing disease resistance protein	*
14,353,023	14,355,842	Eucgr.F01021	RPS5-like 1	*
14,412,056	14,417,904	Eucgr.F01022	NB-ARC domain-containing disease resistance protein	*
14,641,958	14,644,123	Eucgr.F01053	NB-ARC domain-containing disease resistance protein	*
14,657,332	14,660,442	Eucgr.F01056	S-locus lectin protein kinase family protein	*
14,657,332	14,660,442	Eucgr.F01056	S-locus lectin protein kinase family protein	*
14,696,277	14,701,400	Eucgr.F01059	NB-ARC domain-containing disease resistance protein	*
14,815,223	14,816,993	Eucgr.F01069	MYB-like 102	*
15,069,585	15,070,985	Eucgr.F01101	Disease resistance protein (CC-NBS-LRR class) family	*
15,210,703	15,213,899	Eucgr.F01111	RPS5-like 1	*
15,250,695	15,254,201	Eucgr.F01115	RPS5-like 1	*
15,415,646	15,421,399	Eucgr.F01132	NB-ARC domain-containing disease resistance protein	*
15,448,293	15,452,459	Eucgr.F01135	NB-ARC domain-containing disease resistance protein	*

	15,478,485	15,482,610	Eucgr.F01136	NB-ARC domain-containing disease resistance protein	*
	15,534,165	15,538,393	Eucgr.F01139	NB-ARC domain-containing disease resistance protein	*
	15,563,581	15,568,697	Eucgr.F01140	NB-ARC domain-containing disease resistance protein	*
	15,657,955	15,662,148	Eucgr.F01142	NB-ARC domain-containing disease resistance protein	*
	15,865,901	15,866,847	Eucgr.F01164	ethylene responsive element binding factor 4	
	16,626,830	16,627,911	Eucgr.F01254	Disease resistance protein (TIR-NBS-LRR class) family	
	16,763,204	16,766,148	Eucgr.F01261	disease resistance protein (TIR-NBS-LRR class), putative	
	17,093,364	17,096,673	Eucgr.F01276	Disease resistance protein (TIR-NBS-LRR class) family	
	17,184,416	17,186,268	Eucgr.F01287	disease resistance protein (TIR-NBS-LRR class), putative	
	17,250,275	17,253,833	Eucgr.F01291	disease resistance protein (TIR-NBS-LRR class), putative	
7	32,147,370	32,148,942	Eucgr.G01637	Peroxidase superfamily protein	
	32,194,566	32,196,256	Eucgr.G01642	Peroxidase superfamily protein	
	32,266,401	32,266,880	Eucgr.G01645	basic leucine-zipper 44	
	34,175,528	34,177,387	Eucgr.G01774	myb domain protein 4	
	34,466,879	34,470,789	Eucgr.G01790	LRR and NB-ARC domains-containing disease resistance protein	
	34,466,879	34,470,789	Eucgr.G01790	NB-ARC domain-containing disease resistance protein	
	35,397,622	35,402,385	Eucgr.G01856	Disease resistance protein (CC-NBS-LRR class) family	
	35,410,379	35,413,684	Eucgr.G01857	Disease resistance protein (CC-NBS-LRR class) family	
	37,103,269	37,106,472	Eucgr.H05030	Disease resistance protein (CC-NBS-LRR class) family	
	38,055,682	38,059,232	Eucgr.G01970	related to AP2 6l	
	38,919,577	38,922,090	Eucgr.G02007	NB-ARC domain-containing disease resistance protein	
	39,663,507	39,665,156	Eucgr.G02036	1-cysteine peroxiredoxin 1	
	41,596,187	41,598,338	Eucgr.G02167	Polyketide cyclase/dehydrase and lipid transport superfamily protein	*
	41,656,072	41,656,602	Eucgr.G02169	MLP-like protein 28	*
	41,682,857	41,683,437	Eucgr.G02170	Polyketide cyclase/dehydrase and lipid transport superfamily protein	*
	41,686,749	41,687,696	Eucgr.G02171	Polyketide cyclase/dehydrase and lipid transport superfamily protein	*
	41,692,018	41,692,893	Eucgr.G02172	Polyketide cyclase/dehydrase and lipid transport superfamily protein	*

	41,699,450	41,700,021	Eucgr.G02173	Polyketide cyclase/dehydrase and lipid transport superfamily protein	*
	41,713,916	41,714,487	Eucgr.G02174	Polyketide cyclase/dehydrase and lipid transport superfamily protein	*
	41,719,490	41,728,355	Eucgr.G02175	Polyketide cyclase/dehydrase and lipid transport superfamily protein	*
	41,760,027	41,760,924	Eucgr.G02176	Polyketide cyclase/dehydrase and lipid transport superfamily protein	*
	41,794,426	41,795,278	Eucgr.G02177	Polyketide cyclase/dehydrase and lipid transport superfamily protein	*
	41,803,825	41,805,274	Eucgr.G02178	MLP-like protein 423	*
	42,032,893	42,034,733	Eucgr.G02195	myb domain protein 4	*
	43,970,276	43,972,603	Eucgr.G02326	Integrase-type DNA-binding superfamily protein	
	44,078,311	44,082,523	Eucgr.G02341	bZIP transcription factor family protein	
	44,864,209	44,866,024	Eucgr.G02397	Peroxidase superfamily protein	
	44,870,816	44,872,336	Eucgr.G02398	Peroxidase superfamily protein	
	45,459,770	45,465,360	Eucgr.G02459	WRKY DNA-binding protein 57	
	45,554,940	45,558,600	Eucgr.G02469	WRKY DNA-binding protein 3	
9	1,717,986	1,722,236	Eucgr.I00084	disease resistance protein (TIR-NBS-LRR class), putative	
	4,436,726	4,440,641	Eucgr.I00210	NB-ARC domain-containing disease resistance protein	
	4,458,378	4,461,233	Eucgr.I00211	NB-ARC domain-containing disease resistance protein	
	4,527,826	4,528,488	Eucgr.I00216	.	
	5,684,772	5,686,536	Eucgr.I00291	Integrase-type DNA-binding superfamily protein	
	5,762,437	5,763,026	Eucgr.I00292	Integrase-type DNA-binding superfamily protein	
	5,900,669	5,904,355	Eucgr.I00305	WRKY DNA-binding protein 13	
	6,044,956	6,048,416	Eucgr.I00316	WRKY DNA-binding protein 51	
	6,069,028	6,073,301	Eucgr.I00317	WRKY DNA-binding protein 51	
	6,125,999	6,130,640	Eucgr.I00320	.	
	8,172,780	8,174,600	Eucgr.I00422	Integrase-type DNA-binding superfamily protein	
	8,824,106	8,825,662	Eucgr.I00442	myb domain protein 92	
	9,552,903	9,553,865	Eucgr.I00469	receptor like protein 7	
	9,628,977	9,632,256	Eucgr.I00474	LRR and NB-ARC domains-containing disease resistance protein	

9,736,750	9,741,808	Eucgr.I00481	LRR and NB-ARC domains-containing disease resistance protein	
11,110,974	11,113,901	Eucgr.I00549	LRR and NB-ARC domains-containing disease resistance protein	*
11,376,894	11,381,441	Eucgr.I00564	AINTEGUMENTA-like 6	*
13,901,347	13,902,702	Eucgr.I00663	myb domain protein 84	
14,069,539	14,070,098	Eucgr.I00674	.	
14,834,780	14,836,948	Eucgr.I00717	root hair specific 19	
17,163,851	17,165,254	Eucgr.I00828	MYB-like 102	
18,071,433	18,074,088	Eucgr.I00882	WRKY DNA-binding protein 11	
18,283,207	18,287,258	Eucgr.I00892	Integrase-type DNA-binding superfamily protein	
18,283,207	18,287,258	Eucgr.I00892	Integrase-type DNA-binding superfamily protein	
18,415,003	18,415,875	Eucgr.I00894	MLP-like protein 423	
18,420,452	18,421,950	Eucgr.I00895	.	
18,422,334	18,422,916	Eucgr.I00896	MLP-like protein 423	
18,443,857	18,444,439	Eucgr.I00897	MLP-like protein 423	
18,462,115	18,462,697	Eucgr.I00898	MLP-like protein 423	
18,476,165	18,476,996	Eucgr.I00900	MLP-like protein 423	
18,479,951	18,480,517	Eucgr.I00902	MLP-like protein 423	
22,550,148	22,553,714	Eucgr.I01145	abscisic acid responsive elements-binding factor 3	
22,588,805	22,589,862	Eucgr.I01153	redox responsive transcription factor 1	
22,717,283	22,719,174	Eucgr.I01176	Peroxidase superfamily protein	

[a] Linkage group

[b] Gene identification as noted in *Eucalyptus grandis* genome v2

[c] Gene annotation as noted in *Eucalyptus grandis* genome v2

Supp. 3.4 Location of genes differentially expressed in *Eucalyptus grandis* exposed to *Puccinia psidii* from Moon et al. (2007)

LG[a]	Physical position (bp)[b]	GI number[c]	Library[d]	Annotation[e]
3	30,731,089	57898928	S	N-rich protein
7	37,651,792	94420039	S	hypothetical protein
7	42,301,096	50299509	S	calmodulin cam-205
9	27,775,344	7643794	R	ADP-ribosylation factor
9	22,468,154	3023419	R	caffeoyl-CoA O-methyltransferase
9	22,700,395	13430182	S	ribosomal protein L17

[a] Linkage group

[b] Midpoint of marker sequence based on BLAST against *Eucalyptus grandis* genome v2

[c] GenInfo Identifier number

[d] S = overexpressed in susceptible trees, R = overexpressed in resistant trees

[e] Putative identification from Moon et al. (2007)

Supp.3.5 Digenic interacting QTL for rust resistance in the pseudo-testcross maps of the male and female parents of the *Eucalyptus globulus* F_2 population

Map	Trait[a]	LG1[b]	cM1[c]	Adj. marker1[d]	LG2[b]	cM2[c]	Adj. marker2[d]	Add-QTL1[e]	Add-QTL2[f]	Add-Add[g]	LOD[h]	PVE[i]
Male	HR	1	77.688	640651 643645	3	57.484	637279 573637	-0.001	-0.0127	0.0874	10.7379	7.1123
		1	37.688	574251 Embra12	8	50.091	Embra47 569204	-0.0318	-0.0267	0.0743	7.6687	6.716
		2	0	Embra58 Embra63	5	36.193	503594 503419	0.0022	-0.008	0.0705	6.078	4.4029
		2	25	Embra58 Embra63	10	0	565164 504908	-0.0061	0.007	0.0576	3.987	3.1897
		3	2.484	643036 574367	6	3.004	Embra1643 569441	0.005	0.007	0.0807	9.9314	5.973
		3	32.484	638730 640311	10	35	570771 565250	0.0156	0.0039	0.0966	5.9709	8.818
		4	7.174	564631 600106	8	50.091	Embra47 569204	-0.0137	-0.0277	0.0956	11.848	8.9052
		5	36.193	503594 503419	10	30	570771 565250	-0.0022	0.0074	0.0712	4.4872	4.7806
		5	6.193	565220 Embra41	11	45	641615 EMBRA1319	0.009	-0.0116	0.1322	13.4889	16.0057

		6	113.004	575359 Es157	11	45	641615 EMBRA1319	0.0035	-0.0072	0.1039	5.6708	10.043
		10	0	565164 504908	11	90	EMBRA747 639297	0.008	-0.0252	0.0936	6.3394	9.2657
Male	NR	1	22.688	643444 599787	7	20.468	640671 599402	-0.0002	-0.0028	-0.0797	7.8364	16.4725
		1	22.688	643444 599787	9	40	642388 Embra941	-0.0096	-0.0148	0.0768	5.9148	16.1774
		2	70	Embra72 566118	7	45.468	643269 Embra1761	0.0053	-0.0178	0.0355	4.3568	4.0748
		3	97.484	567027 636581	6	18.004	569441 641098	0.0003	-0.0037	0.0385	4.1749	3.8217
		3	32.484	638730 640311	7	60.468	570240 575285	-0.01	-0.0158	-0.0811	11.545	18.2126
		3	72.484	567027 636581	8	85.091	Embra30 503506	-0.0071	0.0023	0.0519	7.7772	7.1256
		5	26.193	503594 503419	11	75	566948 CRC2	0.0096	-0.003	0.0539	9.2662	7.4438
		6	93.004	Embra173 504481	10	55	Embra153 Embra385	-0.0038	0.0063	-0.0457	6.9753	5.5507
		7	20.468	640671	10	0	565164	-0.003	-0.0036	-0.0384	4.6759	3.9819

				599402			504908					
		8	105.091	564050	9	5	640922	-0.0052	0.0046	-0.0589	4.4942	9.0102
				568821			571185					
		8	70.091	573548	11	0	CRC10	0.0019	0.008	0.0455	7.3533	5.3719
				Embra30			566850					
Female	HR	2	0	Embra58	8	8.847	CRC6	-0.018	-0.0097	0.0588	3.8506	3.4698
				564330			Embra47					
		2	85	Embra27	11	85	504103	-0.0086	0.0074	-0.1556	9.816	22.312
				564988			EMBRA747					
		3	100	565405	10	8.126	638769	0.0089	-0.024	0.0752	5.5641	5.9959
				640146			504255					
		6	108.004	569481	9	43.682	Embra941	0.0247	0.0122	-0.0798	6.0271	6.7636
				Es157			EMBRA18					
Female	NR	1	11.899	Eg84	3	50	639640	-0.0025	-0.1024	0.0362	3.6474	31.6051
				Eg65			643885					
		2	60	573002	11	25	570464	-0.0561	0.0266	0.0697	3.6259	22.2299
				570681			Eg99					
		3	50	639640	6	33.004	639232	-0.1034	-0.0043	-0.0369	3.693	29.4437
				643885			Embra627					
		4	75	Embra36	5	0.924	Embra618	0.0022	-0.0067	0.0512	6.5565	7.2713
				565463			572492					

4	60	570780	7	90.468	504000	0.0045	0.0221	-0.0762	3.8776	17.8968
		574628			574182					
4	30	503571	9	58.682	EMBRA18	-0.0038	0.0168	0.0392	3.7722	4.675
		Es54			572030					
6	33.004	639232	9	43.682	Embra941	-0.0036	0.0014	0.0397	4.2457	4.0146
		Embra627			EMBRA18					
6	98.004	571688	11	25	570464	-0.0393	0.0279	0.084	4.0514	23.6042
		562934			Eg99					
7	45.468	636589	11	25	570464	0.0484	0.0226	-0.0839	4.3714	24.9979
		571000			Eg99					
8	98.847	570889	11	25	570464	0.0379	0.0194	-0.0896	3.7351	23.6442
		567585			Eg99					
9	8.682	572089	11	25	570464	-0.0435	0.0195	0.0919	4.8188	25.8313
		504674			Eg99					

a NR = No reaction, HR = Hypersensitive reaction
b Linkage group
c QTL LOD peak position
d Adjacent marker to QTL LOD peak
e Genetic additive effect of the QTL1
f Genetic additive effect of the QTL2
g Genetic additive by additive effect from two-dimensional scanning
h LOD score for genetic additive by additive effect from two-dimensional scanning
i The percent variation in resistance score explained for each QTL pair. Pairings were independently tested, and are not additive.

Chapter 4 - A comparison of QTL for resistance to the native pathogen *Quambalaria pitereka* and the exotic pathogen *Austropuccinia psidii* in *Corymbia*

Introduction

The incursion of exotic pathogens is increasing globally (Ennos 2014; Burgess and Wingfield 2016). Correspondingly, the comparison of resistance to exotic and co-evolved pathogens is an emerging field of study aimed at improved understanding of the mechanics of pathogen resistance (Van der Colff et al. 2017, Freeman et al. submitted). Genetic variation in resistance to native pathogens is often maintained in wild plant populations due to spatial and temporal fluctuations in selection pressure, such as favourable environmental conditions for pathogen sporulation and infection (Burdon et al. 2006). In contrast, plants are often protected against newly-encountered exotic pathogens, as their lack of past exposure means the pathogen lacks the necessary genetic mechanisms to overcome innate plant defences (Heath 2000b). However, in cases where exotic pathogens are able to become established, the effects can be devastating due to a lack of co-evolved resistance in naïve plant species. Long-lived forest trees, which generally evolve slower than grasses and herbs (Lanfear et al. 2013), are believed to be especially vulnerable to such incursions, as seen in the cases of chestnut blight (Anagnostakis 1987), sudden oak death (Rizzo and Garbelotto 2003) and dutch elm disease (Karnosky 1979) causing complete population collapse. However, genetic variation for resistance to exotic pathogens has been noted, for example, in white pine challenged by the exotic blister rust (Kinloch and Dupper 2002), and Port Orford cedar when challenged by cedar root disease (Oh et al. 2006).

The mechanisms contributing to non-co-evolved resistance are poorly understood, but the presence of localised cell necrosis in some cases (termed a hyper-sensitive reaction (HR)), suggests a level of pathogen recognition can occur. Direct gene-for-gene recognition of an exotic pathogen, as occurs in the classic avirulence-resistance (AVR/R) gene model (Flor 1942) where the high specificity of pathogen recognition can completely nullify the pathogen effector (Jones and Dangl 2006), generally assumes co-evolution of the host and pathogen. However, a milder hypersensitive response may also be triggered through indirect recognition methods such as molecular ‘decoys’ or ‘guards’ which may occur in co-evolved or non-co-evolved pathosystems (Dangl and Jones 2001; Van der Hoorn and Kamoun 2008).

Myrtle rust is a disease of global importance to Myrtaceae industries, due to its wide host range and potentially destructive effects (Glen et al. 2007; Pegg et al. 2017). Caused by the pathogen *Austropuccinia psidii* (previously *Puccinia psidii*; (Beenken 2017)), this disease results in lesions on actively growing young leaves and shoots, which can lead to defoliation, dieback, malformation and mortality (Ferreira 1983; Minchinton et al. 2014; Pegg et al. 2014b). Native to central and south America (Glen et al. 2007; McTaggart et al. 2018), myrtle rust is now distributed globally, with different strains infecting Myrtaceous flora across many continents (Coutinho et al. 1998; Uchida et al. 2006; Kawanishi et al. 2009; Zhuang and Wei 2011; Roux et al. 2013; Balmelli et al. 2014; McTaggart et al. 2018). The 'pandemic' strain of *A. psidii* was detected in Australia in 2010 in a cut flower farm in New South Wales (Carnegie et al. 2010), and has since spread to several other states (Queensland and Victoria (Pegg et al. 2012), Tasmania (DPIPWE 2018) and the Northern Territory (Westaway 2016)). It has currently been observed on 347 host species from 57 genera of Myrtaceae*,* which is very broad in comparison to other fungal pathogens (Giblin and Carnegie 2014). However, as there are approximately 2,250 Myrtaceous species in Australia (Berthon et al. 2018) the susceptibility of most native species remain untested, and of those assessed detailed knowledge of the intraspecific variation in susceptibility is available for only a few (Zauza et al. 2010; Morin et al. 2012; Tobias et al. 2015; Westaway 2016; Berthon et al. 2018).

A major QTL for resistance to a strain of *A. psidii* present in Brazil was found in *Eucalyptus grandis*, called *PPR1* (Junghans et al. 2003a). Further examination of *Eucalyptus* hybrids revealed a QTL in the same genomic region with a reduced effect, along with several di-genic interacting QTL spread throughout the genome which were also involved in the disease response (Alves et al. 2012). In *E. globulus,* a similar number of di-genic interacting QTL were discovered following artificial inoculation with the pandemic strain present in Australia, along with several highly significant QTL, none of which overlap with *PPR1* (Butler et al. 2016). Several SNPs and candidate genes within eucalypt species have also been associated with resistance to *A. psidii* in Australia (Dillon et al. 2012; Thumma et al. 2013), alluding to the complex and likely polygenic control of resistance to the exotic rust. It is possible that much of the variation in resistance observed in naïve hosts is influenced by constitutive 'preformed' traits such as cuticle thickness or wax content. For instance, Hsieh et al. (2017) argued that overall changes in global gene expression after *A. psidii* inoculation in highly resistant *M. quinquenervia* individuals were small enough to suggest that this resistance may largely reflect preformed host defences. In *E. globulus,* the percent of phenotypic variation explained (PVE) by a major QTL influencing a 'no response' phenotype for rust resistance was higher than that of a major hypersensitive response QTL, supporting

this theory (Butler et al. 2016). However, as noted, hypersensitive responses to the exotic rust were observed in *E. globulus* and other Myrtaceae including *E. pellita* (Santos et al. 2014), *Syzygium luehmannii* (Tobias et al. 2017) and *Melaleuca quinquenervia* (Hsieh et al. 2017), which is consistent with non-specific pathogen recognition triggering hypersensitivity.

As the strength of a hypersensitive response is influenced by the specificity of binding of resistance molecules to pathogen effectors (Hörger et al. 2012), the hypersensitive response to *A. psidii* (when present) will likely be of lower efficacy compared to that for native pathogens with a history of co-evolution with eucalypts; in which highly specific direct pathogen recognition is more likely to occur. This would suggests that in contrast to the relatively polygenic, weaker resistance response to *A. psidii*, native pathogen resistance will be more oligogenic, with the possibility of major effect R genes to combat specific AVR genes. Consistent with this theory, QTL of large effect (verified across families) were detected for resistance to the co-evolved *Teratosphaeria* spp. in an outcross F_2 pedigree of *E. globulus* (Freeman et al. 2008b), along with loci with smaller phenotypic effects from association studies (Thumma et al. 2017).

Quambalaria shoot blight (QSB) is a disease caused by fungal pathogens from the genus *Quambalaria*. Five different species have been described (Pegg et al. 2008), all of which target Myrtaceous plants (Wingfield et al. 1993). *Quambalaria* spp. infect immature leaves and growing shoots, causing damage and distortion which can impact growth and timber quality (Johnson et al. 2009; Brawner et al. 2011). However, in contrast to myrtle rust, QSB is caused by pathogens that have evolved with the Myrtaceae native to Australia (Paap et al. 2008). Another difference between the pathogens is the nature of their infection. *Quambalaria* spp. enter plants via the stomata with no intracellular growth, while *A. psidii* directly penetrates the plant cuticle and produces haustoria within the cells (Xavier et al. 2001; Pegg et al. 2009), suggesting possible differences in the mode of preformed host defences effective against the different pathogens. QSB has caused severe damage to eucalypt field trials and plantations both within Australia (Stone et al. 1998; Carnegie 2007; Lee 2007; Lawson et al. 2008) and overseas (Roux et al. 2006; Zhou et al. 2007; Pegg et al. 2009), threatening their commercial viability.

Eucalypts, of the family Myrtaceae, are a group of trees containing the genera *Angophora, Corymbia* and *Eucalyptus* (Slee et al. 2006). There are over 900 different species of eucalypts spanning 10 subgenera of *Eucalyptus* and two subgenera of *Corymbia* (Brooker 2000). *Corymbia*, only recently classified as a separate genus to *Eucalyptus* (Hill and Johnson 1995), includes 113 species mostly endemic to the tropics, arid and semi-arid

zones of northern Australia (Hill and Johnson 1995; Parra-O. et al. 2009). Of these, *Corymbia citriodora* subsp. *variegata* (spotted gum) is a species with a prominent role in forestry both in Australia and overseas (Rockwood et al. 2008), where it is used for products including timber, charcoal and essential oil (Asante et al. 2001; Lee 2007; Rockwood et al. 2008). *Corymbia citriodora* subsp. *variegata* can readily hybridize with *C. torelliana* (Wallace and Trueman 1995; Wallace et al. 2008), and these hybrids have begun to be employed in plantations due to the increased growth rate and lower susceptibility to QSB inherited from *C. torelliana* (Barbour et al. 2008; Lee et al. 2009; Dickinson et al. 2010). Since the adoption of *Corymbia* as a plantation species, resistance to QSB has been one of the key factors targeted for genetic improvement in Australian breeding programs (Pegg et al. 2011). *Quambalaria pitereka* (previously known as *Ramularia pitereka* (Simpson 2000)) is the main *Quambalaria* species associated with the *Corymbia* complex with 10 known haplotypes (Pegg et al. 2008), and has been widely studied due to its impact on commercial plantations (Dickinson et al. 2004; Paap et al. 2008; Brawner et al. 2011; Pegg et al. 2011). To date, most studies have focused on assessing the variation in QSB (and rust) resistance found within and between provenances of eucalypts across Australia, particularly within the spotted gums (Lee 2007; Pegg et al. 2008; Johnson et al. 2009; Pegg et al. 2009; Brawner et al. 2011; Pegg et al. 2011; Pegg et al. 2014a), but as yet no studies have attempted to discover QTL underlying resistance to these pathogens in *Corymbia*.

This study aims to dissect the genetic basis of variation in myrtle rust and QSB resistance in *C. citriodora* subsp. *variegata* and *C. torelliana* through determination of the number, location and magnitude of effects of QTL underlying this variation. We hypothesise that QTL of larger effect will underlie variation in resistance to the native than the exotic pathogen, due to the potential for direct AVR/R gene systems with a large effect on resistance to the native pathogen. Based on the lack of additive genetic correlation between resistance to these diseases in common germplasm of *C. citriodora* subsp. *variegata* (Freeman et al. submitted), we also hypothesise that there should be very few co-located QTL influencing resistance to the exotic and native pathogens. In order to examine the genetic architecture of resistance more broadly, we also compared our QTL with those found in different eucalypt species, including QTL for other native fungal pathogens (Freeman et al. 2008b) and previously discovered QTL for myrtle rust (Junghans et al. 2003a; Alves et al. 2012; Butler et al. 2016).

Material and methods

Genetic material, inoculation and symptoms assessment

Disease symptoms were assessed in two *Corymbia citriodora* subsp. *variegata* (CCV) x *Corymbia torelliana* (CT) F_1 hybrid pedigrees (360 seedlings in total), resulting from a cross of the same CCV pollen parent (1CCV2-054) with two different CT parents (1CT2-018 and 1CT2-050, Figure 4.1). Seedlings were arranged into trays each containing 12 seedlings of the same pedigree (except for one tray that contained excess individuals of both pedigrees). The trays were arranged into two blocks, one with an approximately even split of each pedigree, and the other containing mostly the larger pedigree (along with the excess individual tray). Initial inoculation was performed when seedlings were three months old. Subsequent inoculations were performed when the seedlings had regrown enough foliage after hedging (see below) resulting in slightly different plant ages between inoculations, with the last inoculation performed at approximately six months of age.

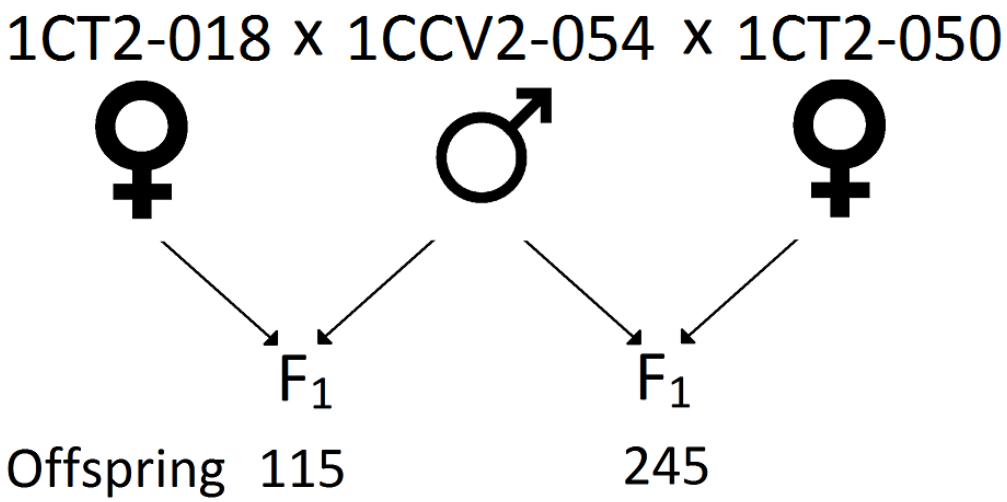


Fig. 4.1 Pedigree of the two *Corymbia citriodora* subsp. *variegata* (CCV) x *Corymbia torelliana* (CT) F_1 hybrid families used for QTL analysis.

Inoculation of CCV x CT hybrid seedlings with *Austropuccinia psidii* was performed as described by Pegg et al. (2014a). In brief, urediniospores previously collected and stored were thawed and added to sterile distilled water with Tween 20 added at a rate of two drops per 100 mL of water. Spore counts were conducted with a haemocytometer and the suspension was adjusted to a concentration of 1×10^5 spores mL^{-1}. Seedlings were inoculated using a fine mist spray applied to both the upper and lower leaf surfaces. To maintain high humidity and leaf wetness the seedlings were covered with a plastic sheet and placed in a controlled environment room at 18 - 20 °C in the dark for 24 hours. Plants

were then moved to a shade-house with disease symptom progression monitored daily. Only seedlings which were actively growing and had new shoots and leaves at inoculation were assessed. Seedlings were assessed 20 days after inoculation for the severity of infection on new shoots and leaves, after which seedlings were cut back to a uniform size, and all diseased tissue removed. Seedlings were allowed to regrow to their original height before separate inoculations with two different haplotypes of *Quambalaria pitereka* (Pegg et al. 2008).

Inoculation of seedlings with *Q. pitereka* was performed as described by Pegg et al. (2009). In brief, the Brisbane Plant Pathology (BRIP) storage collection supplied two isolates of *Q. pitereka* (BRIP 48349, collected from a *Corymbia* hybrid in Mareeba, Queensland; and BRIP 48385, collected from a *C. citriodora* subsp. *variegata* in Beaudesert, Queensland) which were grown on potato dextrose agar for two to three weeks in the dark at 25 °C. A spore suspension (1×10^6 spores mL^{-1}) was obtained by washing plates with sterile distilled water with two drops of Tween 20. Inoculation, assessment and plant cutting back proceeded as above, with inoculation of *Q. pitereka* isolate BRIP 48349 initially, and BRIP 48385 after another round of assessment, cutting back and regrowth.

Following Butler et al. (2016) (Chapter 3), seedlings were assessed for both diseases using a rating scale modified from Junghans et al. (2003b): 1 = no symptoms evident or presence of chlorotic flecking; 2 = presence of a hypersensitive reaction (HR) with fleck or necrosis; 3 = small pustules, < 0.8 mm diameter, with one or two uredinia in rust (or conidiophores bearing conidia in QSB) ; 4 = medium-sized pustules, 0.8 – 1.6 mm diameter with about 12 uredinia/conidiophores; 5 = large pustules, > 1.6 mm diameter, with 20 or more uredinia/conidiophores on leaves, petioles and/or shoots (Pegg et al. 2014a). Leaf shedding was also observed in four cases of QSB infection (two in each pedigree), and was scored as a hypersensitive reaction (Patharkar et al. 2017). Seedlings were then assigned to categories (separately for rust and each QSB haplotype) of: no response (NR, disease score of 1); hypersensitive reaction (HR, disease score of 2); and presence of infection/pustules (INF, disease score 3 - 5). These categories were used to convert the raw disease scores to two different binary scorings for QTL analysis (see below).

QTL analysis

The linkage maps described by Butler et al. (2017b) (Chapter 2) were used for QTL analyses. Due to computational limitations when using dominant markers (Van Ooijen 2009), the smaller “bin maps” described in Butler et al. (2017b) were used. These bin maps were specific to each parent and pedigree, so the QTL analyses were performed using four

linkage maps, one for each CT individual and two for the common CCV parent. The numbering and orientation of linkage groups correspond to the 11 chromosome scaffolds of the *E. grandis* reference genome (Myburg et al. 2014).

QTL analyses were performed using MapQTL 6.0 (Van Ooijen 2009) following the procedure outlined in Butler et al. (2016) (Chapter 3). Putative QTL were declared at two different levels, significant (genome-wide type I error rate < 0.05) and suggestive (chromosome-wide type I error rate < 0.05). The data were initially analysed for QTL using the 1 - 5 disease rating score. To be comparable to the QTL observed in Butler et al. (2016), the data were also analysed using the categories of NR, HR and INF explained above, both for the presence of "no response" QTL through the comparison of NR *versus* HR/INF seedlings, and "hypersensitive response" QTL through the comparison of HR *versus* INF seedlings. The latter analysis required NR seedlings to be excluded from the pedigrees, reducing the sample size and therefore the power to detect QTL. While some suggestive QTL were no longer present in these divided analyses, the discovery of several new QTL in both divisions (many significant) and the increase in LOD score of the remaining QTL gave us confidence to accept these results over the initial analysis based on the original disease rating score (with two exceptions, see Results). To account for possible induced resistance effects caused by the initial rust inoculation, the above analyses were repeated with rust score as a covariate during each QSB h analysis. This caused the removal of three suggestive QTL, but as this also introduced new significant QTL and increased the LOD of the remaining QTL, these results were accepted for the QSB QTL analysis. Due to the reduced size of the CCV54xCT18 pedigree in the comparison of HR *versus* INF seedlings, QSB1 was unable to be used as a covariate, and no test of induced effects was possible.

Positioning of QTL for other pathogens and candidate genes

Several QTL have been previously identified in *E. grandis* and *E. globulus* for resistance to *A. psidii* (Junghans et al. 2003a; Alves et al. 2012) and for other native pathogens, including *Teratosphaeria* (*QMYCO1 - 8*) and a pathogen causing '*Sonderhenia* like-symptoms' (*SOSR*) (Freeman et al. 2008b; Butler et al. 2016). These QTL (along with the QTL detected in this study) were positioned on the *C. citriodora* subsp. *variegata* reference genome (Shepherd et al. 2016; Healey et al. 2017) by the sequence of the marker closest to the QTL peak which yielded an unambiguous hit in the genome (with support from another close marker required when placing via non-peak markers). The likelihood of a pair of QTL co-locating by chance was assessed using the package 'qtlplots' (Rae et al. 2009) in R (R Core Team 2017). Significance of non-random co-location was determined based on 10,000 permutations of

peak QTL locations across all loci, with regions of the genome declared as QTL hotspots when finding a greater number of QTL than would be expected at random (Gosney 2017).

In order to identify positional candidate genes underlying the putative QTL in this study, automatically annotated gene models near each *Corymbia* QTL (up to 2 MB from the QTL peak) were extracted from the *C. citriodora* subsp. *variegata* genome assembly (v1.1, Healey et al. 2017). As these genes currently have no functional annotation, the listed function for the best BLAST hit (< $1e^{-30}$) to the *E. grandis* genome (v2, Myburg et al. 2014) was used as a proxy. These lists of genes were examined specifically for classes of genes previously implicated in resistance (Thumma et al. 2013; Tobias et al. 2017). Due to their suggested involvement with generalised stress responses (Zhang et al. 2006; Cui et al. 2014), as well as influencing the nucleotide-binding leucine-rich-repeat family of genes (NLRs) that are often implicated in resistance (Zhai et al. 2011), microRNA (miRNA) precursors from subfamilies 156, 157 and 172 were specifically searched for in the *C. citriodora* subsp. *variegata* reference genome via BLAST (accepting hits < $1e^{-30}$) using the sequence of previously annotated miRNA in *E. globulus* (Hudson et al. 2014).

Results

Variation in rust and QSB resistance

Between 70 – 105 seedlings were phenotyped for resistance to pathogens in pedigree CCV54xCT18, while between 176 – 220 seedlings were phenotyped in pedigree CCV54xCT50 (Table 4.1). Within each family, seedlings were classified into each resistance class (NR, HR and INF) for each pathogen/haplotype. There were high levels of phenotypic variation observed in resistance to rust in the *Corymbia* mapping crosses but most seedlings were in the susceptible (INF) class. In contrast, both *Corymbia* mapping pedigrees showed relatively high resistance (NR and HR) to both QSB haplotypes (Table 4.1). Resistance scores were not well correlated between inoculations with different pathogens, with only a weak positive phenotypic correlation between rust and QSB1 in pedigree CCV54xCT50 (Spearmans $r_{186} = 0.16$, $P < 0.05$), and a weak negative correlation between QSB1 and QSB2 in the smaller pedigree CCV54xCT18 (Spearmans $r_{56} = -0.29$, $P < 0.05$). When both pedigrees were combined, there was no correlation detected between QSB1 and QSB2 (Spearmans $r_{221} = 0.01$, $P > 0.05$, Figure 4.1).

Table 4.1 Summary of disease ratings from myrtle rust (Rust) and two haplotypes of Quambalaria shoot blight (QSB1 & QSB2) inoculations of the *Corymbia citriodora* subsp. *variegata* (CCV) x *C. torelliana* (CT) F_1 hybrid pedigrees

Inoculation	Cross	No. Seedlings	Mean resistance (SD)	No. NR[a]	No. HR[a]	No. INF[a]
Rust	CCV54xCT18	105	2.8 (± 1.4)	28	21	56
	CCV54xCT50	220	2.7 (± 1.1)	29	73	118
QSB1	CCV54xCT18	70	1.6 (± 0.8)	40	19	11
	CCV54xCT50	196	1.4 (± 0.6)	133	48	15
QSB2	CCV54xCT18	81	1.5 (± 0.8)	53	20	8
	CCV54xCT50	176	2.0 (± 1.1)	68	64	44

[a] NR = no response, disease rating of 1; HR = hypersensitive response, disease rating of 2; INF = infection symptoms, disease rating of 3, 4 and 5

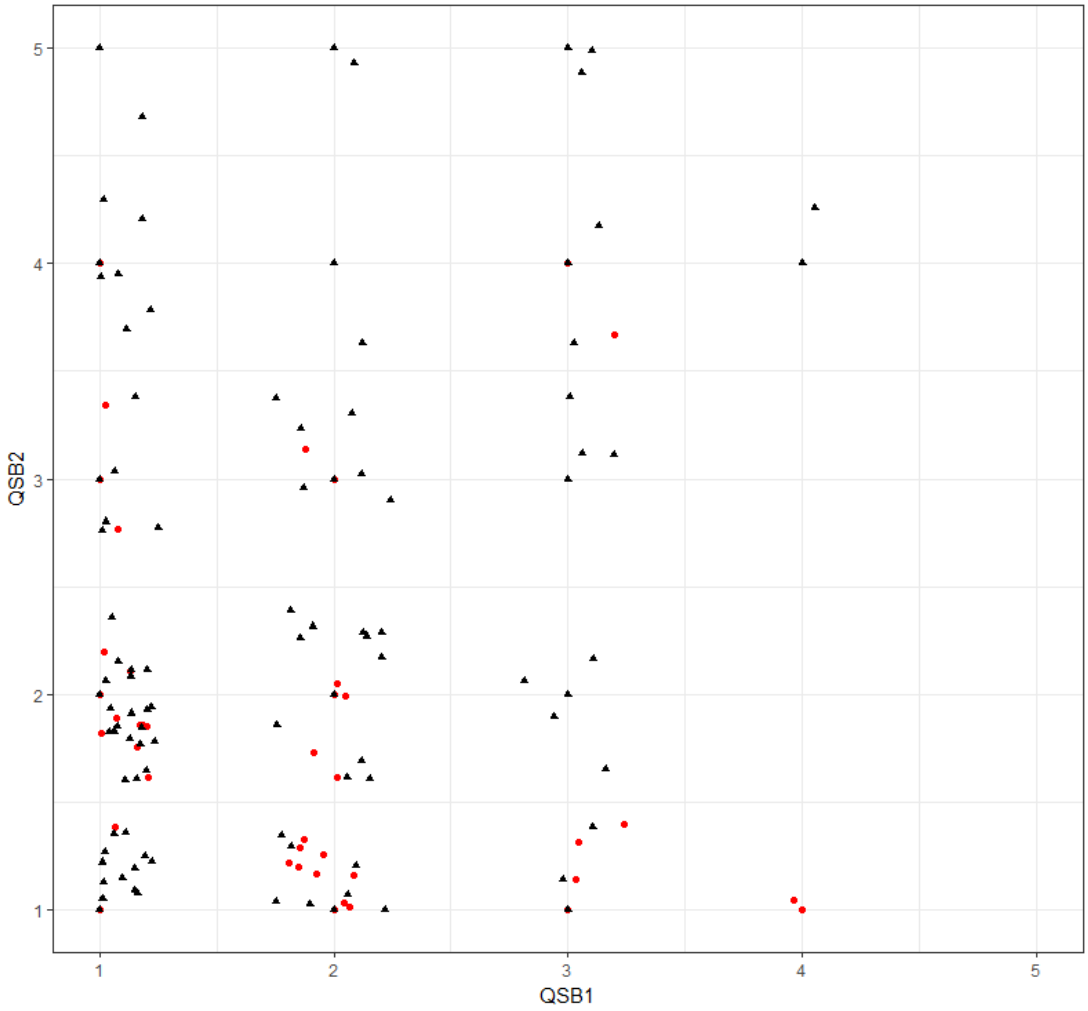


Fig. 4.1 Disease scores of seedlings in both *Corymbia* pedigrees when inoculated with two different haplotypes of Quambalaria shoot blight (QSB1 and QSB2). Pedigree CCV54xCT18 is indicated by red circles, while pedigree CCV54xCT50 is indicated by black triangles. As disease scores are always discrete, a small amount of random variation is added to scores to allow all points to be seen. No significant correlation between QSB1 and QSB2 resistance scores within all individuals was detected (although a weak negative correlation was detected in the smaller pedigree CCV54xCT18).

QTL discovery and effect size

Across the three pathogens and four linkage maps, 11 QTL were found at the significant level and nine at the suggestive level (Table 4.2, Supp. 4.2). Of these, six were associated with rust resistance (of which five segregated from *C. torelliana*), nine were associated with resistance to QSB1, and five were associated with resistance to QSB2. A threshold of 5 cM between QTL peaks was used to determine co-location, which corresponds to 2 MB in the *Corymbia* reference genome, on average. There were two occurrences of co-located QTL for QSB2 resistance from both the NR and HR response categories, so these QTL were categorized as "general" (GEN) and presented as a single QTL based on the raw disease severity scores. No other co-locating QTL within the linkage maps were detected for the same disease inoculation. Of these 20 QTL, 10 were found specifically for no response (NR), eight for the hypersensitive response (HR) and two for a general response (GEN), with all pathogens and haplotypes exhibiting both responses.

Table 4.2 Putative QTL underlying variation in myrtle rust and Quambalaria shoot blight resistance in the two *Corymbia* F_1 hybrid pedigrees

Linkage map[a]	Disease	Category[b]	LG[c]	cM[d]	Adj. marker[e]	Flanking markers[f]	LOD[g]	PVE[h]	Physical position (bp)[i]	
									Peak	Flanking
CCV54 (xCT50)	Rust	HR	9	38.5	10277690	10353587 10362662	4.69**	10.7	16,660,128	15,211,998 18,265,765
	QSB1	NR	8	10.1	10339561[d]	10370202 10313232	4.33*	9.7	8,610,406	5,716,670 14,410,825
	QSB2	GEN	8	99.9	10279281	10312744 10572101[e]	8.63***	21.1	45,835,233	33,153,261 48,188,667
	QSB2	NR	2	120.1	10323977	10327289 10327704	3.10	7.5	37,842,033	34,428,366 42,532,134
	QSB2	HR	1	103.7	10314561[d]	10284185[e] 10313889	3.45**	11.1	29,631,451	25,744,178 30,496,956
CT50	Rust	NR	6	87.0	10315746	10316739 10280763	3.71*	7.2	31,342,621	28,098,405 37,405,435
	Rust	NR	4	39.3	10356168[d]	10572286 10339836	2.33	4.5	17,873,452	7,126,279 24,228,842
	QSB1	NR	10	75.8	10323463	10280841 10280468	2.81	6.2	20,810,948	17,391,631 23,380,806
	QSB1	HR	11	78.7	10313244[d]	10573908 10315161[e]	5.31***	34.5	3,753,893	4,685,749 3,251,212
	QSB1	HR	6	35.0	10290812	10337293 10288059	2.77	15.9	9,473,080	3,658,448 20,502,979
	QSB2	GEN	4	0	10341479	End 100042813	3.58*	9.3	178,254	End 6,003,090
CCV54 (xCT18)	QSB1	NR	2	19.1	10369199[d]	10355448 10377227	5.06**	22.4	6,674,313	4,684,847 6,816,460
	QSB1	NR	8	33.8	10339561	10573001[e] 10353386	4.23*	18.2	8,610,406	5,303,086 4,481,070
	QSB1	NR	7	38.2	10359541	10277174[e]	2.29	9.2	9,413,339	2,117,963

						10339075				36,341,060
CT18	Rust	NR	6	89.9	10573687	10325045	7.62***	26.1	25,789,255	21,734,342
						10341636				26,283,338
	Rust	NR	10	25.5	10568868	10570318	2.89	9.0	10,257,470	1,818,074
						10326400				7,681,296
	Rust	HR	3	48.3	10280956	10274514	2.81	15.5	16,826,863	6,293,383
						10357059				24,604,908
	QSB1	HR	3	111.5	10280176	10285446	3.29	22.6	41,218,655	38,322,137
						10279708				42,465,048
	QSB1	HR	6	115.2	10305288	10280042	5.29*	44.4	32,317,102	32,011,033
						10276775[e]				35,630,178
	QSB2	HR	7	113.7	10575033	10349358	3.85	48.6	37,571,582	36,887,714
						End				End

[a] Four linkage maps were used to represent the parents of each pedigree, including the common parent CCV54

[b] NR = No response, HR = Hypersensitive reaction, GEN = Generalised response (overlapping QTL for NR and HR within one linkage map)

[c] Linkage group

[d] QTL LOD peak position

[e] Adjacent marker to QTL LOD peak. For those markers flagged the sequence of the peak marker was located on a scaffold in the *Corymbia* reference genome incongruent with the linkage map, so the closest adjacent marker anchored to the correct scaffold is reported. The sequence of these markers are presented in Supp. 4.1.

[f] Markers flanking outside of 2 LOD confidence interval from QTL peak. See above for explanation of flagged markers.

[g] Peak LOD score for each QTL. Genome-wide significance is indicated by *** P < 0.001, ** P < 0.01, * P < 0.05. The remaining QTL were significant at the suggestive level (chromosome-wide type I error rate < 0.05).

[h] The proportion of phenotypic variation explained by each QTL

[i] Midpoint of marker sequence based on BLAST against *Corymbia* genome v1.1

To compare the effect size of the different QTL for each pathogen/haplotype, the mean and maximum estimated percentage of variation explained (PVE) by the QTL were contrasted within both pedigrees (Table 4.3). The maximum effect sizes of the rust QTL were smaller than that of QTL for QSB1 or QSB2 in both pedigrees, as were the mean effect sizes (although not significantly, in CCV54xCT50 [$F_{2,8} = 0.94$, $P = 0.43$], CCV54xCT18 [$F_{2,6} = 2.78$, $P = 0.14$] or across both pedigrees [$F_{2,17} = 0.82$, $P = 0.46$]).

Table 4.3 Comparison of QTL percentage variation explained (PVE) for each pathogen within pedigrees

Disease	No. QTL		Mean PVE (SE)[a]		Maximum PVE[b]	
	CCV54xCT50	CCV54xCT18	CCV54xCT50	CCV54xCT18	CCV54xCT50	CCV54xCT18
Rust	3	3	7.5 (1.5)	16.9 (4.1)	10.7	26.1
QSB1	4	5	16.6 (5.5)	23.4 (5.2)	34.5	44.4
QSB2	4	1	12.2 (2.6)	48.6 (NA)	21.1	48.6

[a]The average PVE and standard error in parenthesis of all QTL for a specific disease in each pedigree. There was only a single QSB2 QTL in pedigree CCV54xCT18, so no standard error was calculated.
[b]The maximum PVE exhibited by a QTL for a disease in each pedigree.

Co-location of QTL from different crosses using the Corymbia reference genome

All rust and QSB QTL were mapped to the *Corymbia* reference genome via their peak marker sequence. There was one instance of a rust QTL co-locating (within ± 2 MB) with a QSB QTL, occurring on chromosome 6 (specifically *Rust_CT50_NR6* and *QSB1_CT18_HR6,* Figure 4.2). These QTL were from different *C. torelliana* linkage maps and different response categories, but were both found at a significant LOD level (Table 4.2). A QSB1 QTL for the NR response was also found in both male maps in the same location on chromosome 8 (Figure 4.2). The lack of co-location between other resistance QTL in CCV54 is not unexpected, due to the different genetic backgrounds introduced by the separate CT parents.

Several QTL associated with *Sonderhenia* and *A. psidii* resistance in *E. globulus* and *E. grandis* were co-located with the QTL for rust and QSB resistance found in CCV and CT (Figure 4.2). Specifically, *Rust_CT18_HR3* was co-located with both *PPR2* and *SOSR3* on chromosome 3, *QSB1_CT50_HR6* was co-located with both *SOSR4* and *PPR4* on chromosome 6, *QSB1_54x18_NR7* was co-located with *QMYCO3* and *SOSR1* on chromosome 7, and *Rust_CT18_NR10* with *SOSR2* on chromosome 10. The similar location of *QMYCO3* with a QSB QTL is notable, as three other QTL for *Teratosphaeria* resistance were within 5 MB of a QTL for QSB (Figure 4.2). After assessing the likelihood of all QTL co-locating by chance, the clustering of QTL on both chromosomes 6 and 7 were found to be significant ($P < 0.05$, Figure 4.2).

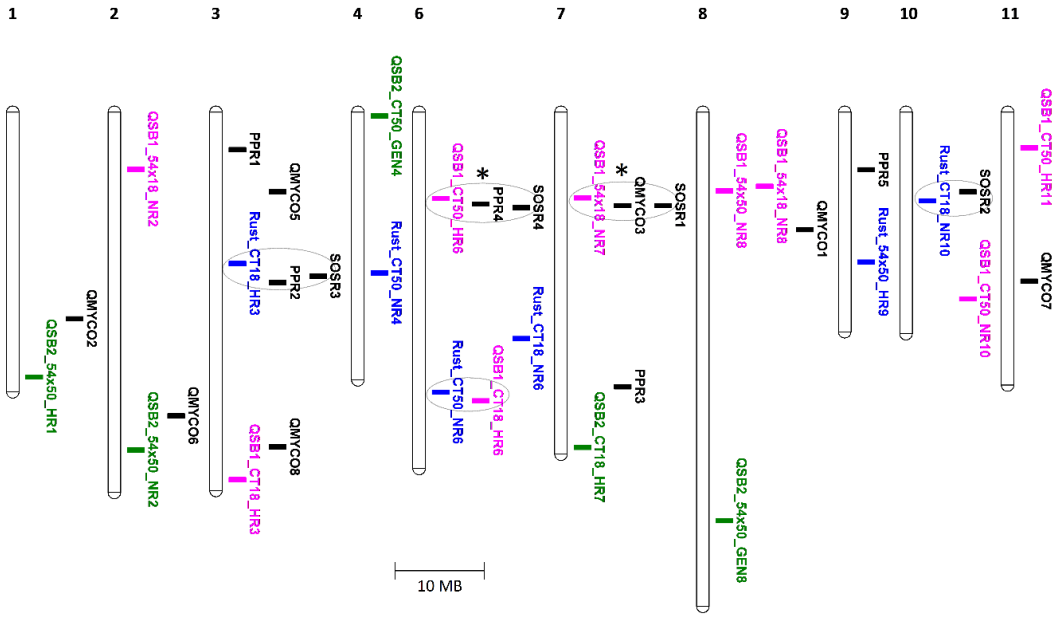


Fig. 4.2 Co-location of eucalypt disease resistance QTL on the *Corymbia* reference genome. Base pair positions of QTL peaks from this study are listed in Table 4.2, while those from *E. globulus* and *E. grandis* are listed in Supp. 4.3. Circled QTL indicate those meeting the criteria for co-location (within ± 2 MB), and those labelled with a * indicate that the association of these QTL is greater than would be expected by chance. Only chromosomes with disease QTL are shown.

Candidate gene discovery

All gene models present within 2 MB of a QTL peak in the *C. citriodora* subsp. *variegata* reference genome (v1.1, Shepherd et al. 2015; Healey et al. 2017) were collated as potential positional candidates. As these models are unannotated, a putative functional annotation was assigned to them based on the best BLAST hit of their sequence to annotated genes on the *E. grandis* reference genome. Under each QTL several classes of genes related to plant resistance responses were found including peroxidases, chitinases, pathogenesis proteins, NLRs, zinc fingers, TIRs and other families of transcription factors including MYB, WRKY and AP2 (Supp. 4.4). Of particular interest were those genes under the co-located QTL positions on chromosomes 3, 6 and 7. Notably, in all three cases the most frequent genes were NLRs, with 9 - 22 of each found within the 2 MB interval. Several microRNA genes were also found to co-locate with QTL in this study, with putative miRNA156, miRNA157 and miRNA172 precursor genes found co-locating with QTL for rust and QSB resistance (Supp. 4.5)

Discussion

This study reports QTL associated with resistance to both the native disease Quambalaria shoot blight (QSB, caused by *Quambalaria pitereka*), and the exotic disease myrtle rust

(caused by *Austropuccinia psidii*) in two hybrid crosses of *Corymbia citriodora* subsp. *variegata* (CCV) x *C. torelliana* (CT). Due to the impact of QSB in commercial plantations it is a key trait targeted for genetic improvement in *Corymbia* breeding programs. These are the first QTL reported for resistance to QSB in eucalypts and, indeed, the first for resistance to any disease in *Corymbia*. Notably, most QTL for rust resistance segregated from *C. torelliana,* consistent with the known lower incidence of infection compared to *C. citriodora* subsp. *variegata* (Pegg et al. 2014a). As we initially hypothesised, i) the comparison of QTL effect sizes supports the theory of relatively oligogenic control of resistance to native pathogens and polygenic control of resistance to exotic pathogens; ii) there was very little co-location between QTL influencing resistance to the exotic and native pathogens within these pedigrees. The lack of co-location between QTL for separate haplotypes of QSB is also of note, suggesting high specificity in these resistance responses. Together, these findings significantly advance our understanding of the genetic architecture underlying variation in host resistance in these pathosystems.

Differing QTL resistance effect size between native and exotic pathogens

Consistent with our hypothesis of oligogenic control of resistance to native pathogens *versus* polygenic control of exotic, both the maximum percentage of variation explained (PVE) and mean PVE was greater for the QSB QTL compared to the rust QTL, both within and between pedigrees. Several studies have alluded to the polygenic control of rust resistance in eucalypts (Alves et al. 2012; Thumma et al. 2013; Butler et al. 2016; Tobias et al. 2017). The QTL influencing the resistance response to rust are likely to underlie a combination of preformed defences as well as indirect pathogen detection systems leading to hypersensitive responses, as has been suggested for other naïve plant/pathogen systems (Jones and Dangl 2006). The presence of a hypersensitive reaction to the exotic pathogen is not unexpected based on previous observations across the Myrtaceae (Tobias et al. 2016). While in pedigree CCV54xCT50 rust QTL effects were roughly even for ‘no response’ (NR) and a hypersensitive response (HR), in pedigree CCV54xCT18 the highest effect QTL in rust influenced NR. In combination with evidence from previous studies (Butler et al. 2016; Hsieh et al. 2017), this suggests that preformed resistance may be one of the more important factors when determining susceptibility to rust.

In addition to the above mechanisms, native pathogens will likely be detected by the host species via direct recognition by specific R genes of relatively large effect due to their shared evolutionary history, which may contribute to a more oligogenic control of resistance (Jones and Dangl 2006; Kushalappa et al. 2016). Large effect size in QTL for native pathogen resistance has been observed previously (Freeman et al. 2008b)),

consistent with this hypothesis. In this study, the QTL with the highest PVE for both haplotypes of QSB were underlying the hypersensitive response (HR), which is consistent with specific R genes rather than multiple indirect or preformed resistance mechanisms. The higher PVE of these QTL compared to those for rust is also consistent with this hypothesis, given that a stronger resistance response due to binding specificity to a co-evolved native pathogen is expected (Hörger et al. 2012). In addition to the higher PVE by these QTL, the lower disease scoring observed for *Q. pitereka* in both pedigrees also supports a stronger resistance to the native pathogen. Further investigation across a broader range of host germplasm and in other pathosystems will be required to test the general applicability of this finding, ideally using controlled inoculation by native and exotic pathogens in common host germplasm.

Notable co-locations between QTL

Our results support the hypothesis that there should be little correlation between QTL underlying resistance to the exotic and native pathogens. Only one QTL associated with rust resistance co-located with a QSB QTL in these pedigrees, found on chromosome 6 in the *Corymbia* reference genome, with both QTL originating from different CT individuals. This result was expected given previous studies have found the genetic correlation between susceptibility to rust and QSB to be non-significant (Freeman et al. submitted). Interestingly, the co-located QTL underlie different resistance responses to each pathogen, the “no response” in rust, and the hypersensitive reaction in QSB. If these different QTL indeed reflect the action of separate mechanisms, potential explanations include a localized array of genes involved in different resistance responses or shared steps in separate resistance pathways (Dangl and Jones 2001; Tobias et al. 2016).

The examination of QTL from other species, however, reveals a great deal more co-location than seen within these *Corymbia* pedigrees. For example, *PPR4* (rust resistance QTL in *E. globulus* (Butler et al. 2016)) co-locates with *QSB1_CT50_HR6* and *SOSR4*, while *SOSR2* co-locates with *Rust_CT18_NR10*. There is also an interesting association (although not reaching the threshold for co-location) between QTL for resistance to both *Q. pitereka* and *Teratosphaeria,* with four of the seven QMYCO QTL peaks within 5 MB of QSB peaks. The association of these resistance QTL provides support for the existence of elements common to resistance responses across pathogens, with potential contributions by arrays of genes influencing common pathogen effector targets (Tobias et al. 2016) or preformed defences.

Another notable co-location was detected for QTL underlying rust resistance in both *Corymbia* and *E. globulus,* positioned (along with a *SOSR* QTL) on chromosome 3. These QTL

are part of an area (spanning 12 MB) which includes the *E. grandis* rust resistance QTL PPR1 (Junghans et al. 2003a), *Teratosphaeria* resistance QTL *QMYCO5* as well as QTL associated with resistance to *Ceratocystis* spp. in *Eucalyptus* hybrids (Rosado et al. 2016), making chromosome 3 one of the more important chromosomes for pathogen resistance in eucalypts. Comparative analysis of sequence using the *E. grandis* reference genome led to the observation that chromosome 3 has undergone almost no rearrangement since diverging from its ancestral eudicot ancestor, a situation also observed in another forest tree, *Populus trichocarpa* (Myburg et al. 2014). Several syntenic sets of genes implicated in disease resistance are present on this chromosome in both genera, suggesting this conservation may be influenced by the advantages provided by the arrangement of these genes, such as coherent regulation and inheritance (Field and Osbourn 2008; Chu et al. 2011).

Positional candidate genes associated with resistance

One of the main categories of genes found under these multiple QTL co-locations are the superfamily of proteins containing nucleotide-binding domains and leucine-rich repeats (NLRs) (Van Der Biezen and Jones 1998). These proteins, which code for most R-genes, interact with pathogen effector molecules to stimulate, either directly or indirectly, an immune response (Cui et al. 2015; Cesari 2017). In *M. quinquenervia* inoculated with *A. psidii,* for instance, up-regulation of NLRs was detected in resistant individuals exhibiting a hypersensitive response, supporting the indirect recognition effects of NLRs or their involvement in that pathway (Hsieh et al. 2017). Given their varied roles, many different selective processes act to diversify this family (Jacob et al. 2013). Often these genes are found in duplicate arrays, having arisen through tandem duplication or transposition/large scale duplication and re-arrangement (Eitas and Dangl 2010). Duplicates are often released from selection and able to evolve new functions, which in the case of pathogen detection can result in the generation of new R genes. *E. grandis* for instance has over 1,200 putative NLR genes with 76% arranged into clusters, indicating a rich library of possible R genes (Christie et al. 2016) and providing one explanation as to the presence of hubs of resistance. The direct binding of multiple unrelated pathogen effectors by a single NLR has also been detected (Cesari et al. 2013), which along with the guard and decoy models of resistance (Dangl and Jones 2001) may contribute to overlapping resistance mechanics for different pathogens (Bisgrove et al. 1994; Kim et al. 2002; Martin et al. 2003).

Another class of genes that were found to collocate with our QTL are the microRNAs (miRNAs, specifically subfamilies 156, 157 and 172). A manual search for precursor loci from these subfamilies revealed several cases of co-location with QTL for resistance to

different pathogens, involving both discrete miRNA loci and arrays (Supp. 4.5). MiRNAs are targeted molecules that cleave mRNA, a post processing step essential for protein formation and regulation (Hobert 2008). As such miRNAs are known to act as regulators in a wide range of functions including plant development, hormone signalling, and stress responses (Zhang et al. 2006). Specifically, the miRNA156/172 family are implicated in both general stress responses including disease resistance (Zhao et al. 2012) and as regulators of the NLR gene families (Zhai et al. 2011). A single miRNA172 precursor was found underlying co-located QTL for resistance to rust and *Sonderhenia,* consistent with a generalised stress response (Supp. 4.5). Co-location between resistance QTLs and these miRNAs in eucalypts has been detected before in *E. globulus* (Butler et al. 2016), further supporting these genes as positional candidates.

Strain specificity of resistance responses

One key finding from this analysis was the lack of co-location between QTL influencing resistance to different haplotypes of QSB in either pedigree. This is in accord with the lack of a significant correlation between disease severity scores for the different haplotypes of *Q. pitereka* in the larger CCV54x50 pedigree (although a weak negative correlation was detected in the smaller pedigree). Haplotype specificity of QTL for resistance is not uncommon; for instance, the response of *Populus* hybrids to infection by different isolates of *Melampsora larici-populina* (causing poplar rust) found that six isolates out of eight were unaffected by a major quantitative resistance factor R_{US} (Dowkiw et al. 2010). This suggests a resistance response that is highly specific, and possibly R gene mediated. If so, this would be consistent with the QTL effect size hypothesis explored earlier. Alternatively, if these QTL are influencing preformed defences such as leaf anatomy (Smith et al. 2017) or chemical composition (Gosney et al. 2016), each haplotype of the pathogen may have specific physical requirements for successful penetration and infection of plant tissue. This has been observed previously in wheat, with different strains of the fungal pathogen *Fusiarium graminearum* observed to penetrate specific tissues with different efficiency, which has been proposed to reflect differences in cell wall thickness between tissues (Jansen et al. 2005).

The lack of co-location between the QTL *PPR1* underlying resistance to a strain of *A. psidii* restricted to Brazil (Junghans et al. 2003a; Mamani et al. 2010), and the resistance QTL discovered in *E. globulus* (Butler et al. 2016) and *Corymbia* for the pandemic strain present in Australia, may indicate strain specific resistance to myrtle rust in the eucalypts or differences in the resistance architecture between different eucalypt species. In support of the latter, SNPs in NLR genes significantly associated with resistance in *E. grandis* to the

pandemic strain of rust were detected near *PPR1* (Thumma et al. 2013) suggesting this locus may be present and not specific to the pathogen strain for which it is was first reported. However, the failure of our QTL studies to detect a QTL for myrtle rust in this genomic location may reflect species specific differences in the genetic architecture underlying resistance, host genomic structural changes such as inter-chromosomal rearrangements (Butler et al. 2017b), or simply a lack of segregation of the homologous loci to *PPR1* in the bi-parental crosses used for our QTL studies. A QTL or association genetic study using the same genetic material inoculated with both the pandemic strains and those strains endemic to Brazil would be ideal to investigate whether resistance to rust is strain specific.

Conclusions

This study reports the first QTL discovered for resistance to fungal pathogens in *Corymbia*. Variation in resistance to the exotic *A. psidii* and the native *Q. pitereka* were influenced by 20 different QTL overall. Notably, QTL for resistance to *Q. pitereka* were generally of larger effect than those for *A. psidii*, which may reflect greater specificity of co-evolved resistance. Only a single instance of co-location between QTL for resistance to these pathogens was detected. However, an examination of the location of these QTL in comparison to those previously discovered for *A. psidii* and additional pathogens in other hosts revealed several instances of co-location, which may reflect common elements influencing resistance responses across pathogens and hosts. These are the first QTL discovered for resistance to Quambalaria shoot blight in any host species, and will be of interest for the further study of this disease, especially in view of the detection of independent QTL influencing resistance to different haplotypes of the same pathogen.

Supplementary material

Supp. 4.1 Peak and adjacent markers to QTL for resistance in *Corymbia*

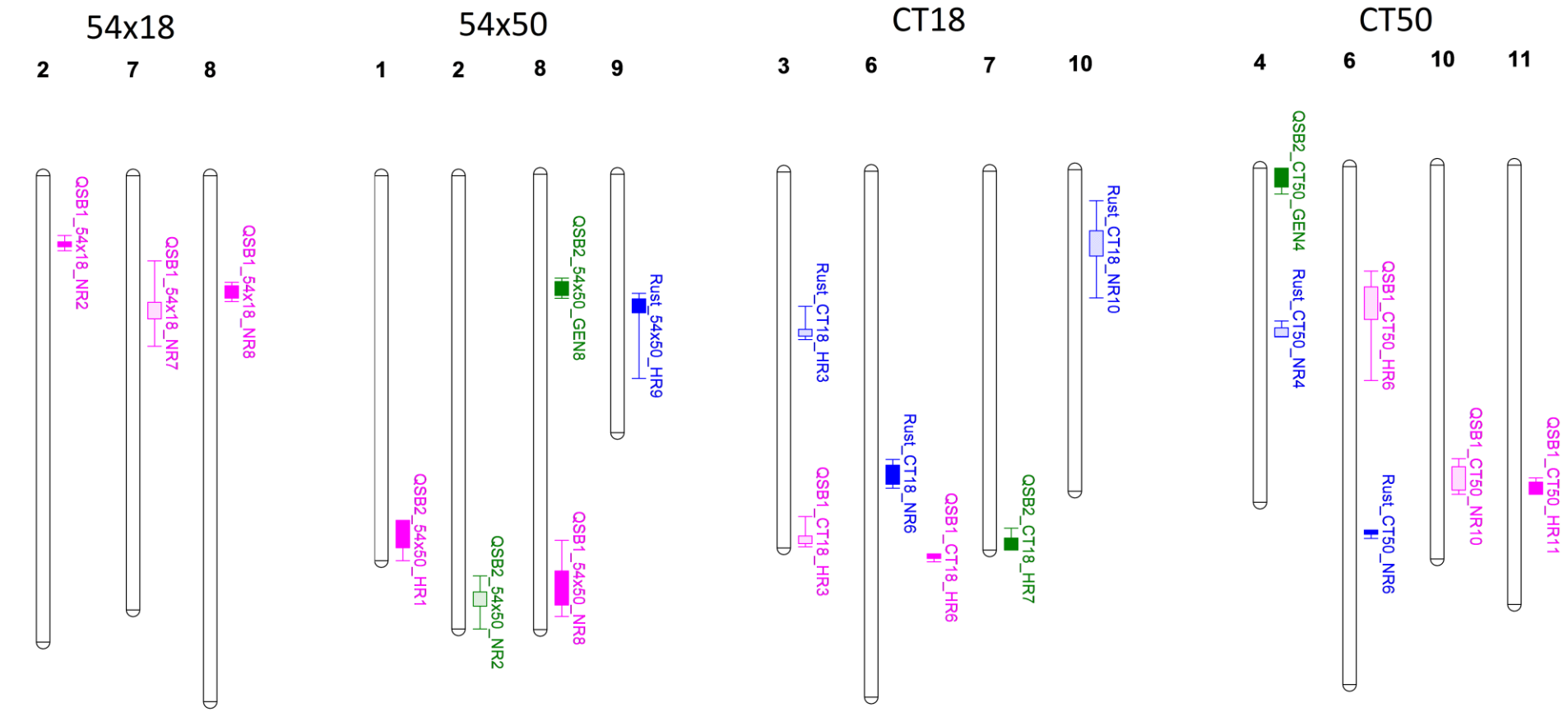
QTL	Adjacent marker[a]	Sequence	Flanking markers	Sequence
Rust_54x50_HR9	10277690	CATGTCCTTACTCTTACTATAGCTTTCTCC TACTACAATTCAATTATTGCAGCACCAGCACCTT	10353587	CATTCTTGTGGGCACTCCTGCTCAGCCCAGCCA TTTGCTATGCCAGGAAGCACAACACTCTGTT
			10362662	TGCACGCAGTTGTTGGCCGTGAGGGCATAAG CGCCGTTGGGCACGAGCAAACCGAGATCGGAAG
QSB1_54x50_NR8	10339561	ACATCTCTATCCCCGCAACTTCAATCCCTC CACAAAAAGAGGGGATGGGGTTTCTCTCGCTTTT	10370202	GTGACTCGACTAGACAGGGATGTCGTGGCTA GGCGACGAATGAAGCTCGATGCTCTGGTCTGGT
	10572274	AATCTTAAGCAATTTTTTTTTTTTTTCCGAG ATCGGAAGAGCGGTTCAGCAGGAATGCCGAGAC	10313232	ATGAAGATGACCTTGGCTGGTTCTCGCTCTGT CTGTTTCTGGGTTTCGCTGGTTGTTCCCGAGA
QSB2_54x50_GEN8	10279281	GTCCCGCGCGTGGAGAAAAAAGGAGCGA AAAGAAAGCGGACCGAGATCGGAAGAGCGGTTCAGC	10312744	TGATGTTATTGGGCCCTGTCAGTCCCCTCGGC CCATCTTCAATCCCTAGAGGACCCTCGGCACG
			10572101	CCGACGGTGTAAAGCGAGCAAAGTACGAGAG AGAGAGAGAGAGAGAGAGATAGAGGTTCACCGA
QSB2_54x50_NR2	10323977	ACGGCTTGTTATGTCTTCCCTGTTTGTTGTT TCTTGTCAAAGCTTTTATTCATAATGGTGGCTC	10327289	TGCGAAAACCCTGCAACCCAAGAGTCCAAAAG AAAATCACCCGTCACCGCCAGTGCACACGAAA
			10327704	TGCCACCTGATGATGGCCATCCACTCTCTGATA CAACAATTGTCAAACTTGACCCACCAACTTT
QSB2_54x50_HR1	10314561	GAGATGCGTCTTCCAAAAGCTGGTCGCTGA TGATGAGCTTGTCCATAGCACCAAGTGTGGGCTG	10284185	AGGGGGGAATGCTAGGACAGATGGACGGACT GACGGAAAGAGAGAGTTGCGAAGAGAAGCTCCG
	10359684	ATCAGGATCGACTAGAATTCCCCACCTCCGCTC TATCAGAAGCCCTTCTACCCAATCGAGTAAG	10313889	AAACAACTTTAAATGCATGTCTAAGTTCGAGCA AAATCATATGCATCGACTGACATGAGAGGGC
Rust_CT50_NR6	10315746	TCAATTTCTCACATGTTAGGTAAGGCAGGGGG CGTGATACGGGGAATTCCCAACTCAGTGCTCT	10316739	CATTGGATTTGATGACATATTTTTACTTCTAATTA CGTCCTTCAGTAGCTCTTGATCTGGATTC
			10280763	TGAGTTCGAAGGTGGCCGAGTTGACCCATCTCTT ATGCTGCCACGTCCCCGAGCGCCGCCCTTT
Rust_CT50_NR4	10356168	GGGCGAAGGACTTAGCGTGCCAGTGAGCTTAG	10572286	ACGAGCTTGAGCTTTCTTGAAGGGGCGGCGGCG

		AAGGAAGTCAATTTATCTTCTCCCAGACATGG		GCGGAGGGTTTCCTCCGAGATCGGAAGAGCG
	10313964	AGTAGAATAAGCATGTTACCTTGGTAAAGCTC TTTCTCTTGATGTAAATTTTAAATCTTGTTAG	10339836	AGGCAGCGACGAGGAAGACGGCGAGAACTCGA GGAAGAAGCTAAGGCTGTCCAAGGATCAGTCG
QSB1_CT50_NR10	10323463	GTTGACTATAGTATGCTGTTATTTACCATCAG GACTGAGATTATTCTCTTTTACTATATTTGCG	10280841	TTAGCTGGGTGCCTGATCAACTGTCTCAGTTTTTA TTGTAGCACCATGATTTGAAATCTTGTGG
			10280468	GGCAAAGATGCTGTCCACATCAACAGAAGACGA GGAGAGAGAGTCATCACCTCTGGAGATAGGT
QSB1_CT50_HR11	10313244	ATGGGATTTCTCCTCCTTTCTCACATGGGGTA GTCTAGCAAGCTGAGCCGAAACCAGTCTCTGA	10573908	TTTCGTGTCATCCTCTAGAACAATTTGCTTCACAA ACTTCAGCAGATCCTACGCAAGCAAGCAA
	10573908	TTTCGTGTCATCCTCTAGAACAATTTGCTTCAC AAACTTCAGCAGATCCTACGCAAGCAAGCAA	10315161	AAGCCACGTTTAGTACCATTCAACCGAACCT CGAAACAAAGTCGACAAAAAGCAATGCAAAAGA
QSB1_CT50_HR6	10290812	AAGCGGCTAAACCAAACTGTTTTCGAAGTTAC ATGGGCATTACACTCTCCCGCCAACCAGTCGC	10337293	AAAGAAACCCAAGGGCTAAGGGAGGAGGAG GCTCTTATCTTAGACGTCCGAGATCGGAAGAGCG
			10288059	TTGGGAGAAAGAAGCGAGCGAAGGATCTCG ACTACTCGAGAGCTTGGCTCAGTGAGTCAGCAGA
QSB2_CT50_GEN4	10341479	GCCCTGGCTCTGTACCACAACGATCTGGAGGG GAGCCTGCCACGTGAGATTGGGATGCTTGGCG	End	
			100042813	GCTGCCTGTTTCCACGCGCGGGAGAAGAAGG GGCGGAAAAAGAAGTCCAAGTGGGCCGAGATCG
QSB1_54x18_NR2	10369199	CCTGATAGACTAAGCTGGCATACACAAGGACG AGGACCAACCGAGATCGGAAGAGCGGTTCAGC	10355448	CATCATATATAGCGGCGGCAATCGAGCACGAT ATTCCTCCCAACATTAACCAACGACAAGAATG
	10355448	CATCATATATAGCGGCGGCAATCGAGCACGAT ATTCCTCCCAACATTAACCAACGACAAGAATG	10377227	TTTCTTCTTCTTCTTTCTTCCTTCTTCTTCTTGCTGC TGTTTTCCGTCGGCCAGAGCTGCTGCT
QSB1_54x18_NR8	10339561	ACATCTCTATCCCCGCAACTTCAATCCCTCCACA AAAAGAGGGGATGGGGTTTCTCTCGCTTTT	10573001	CGTTGGGAGGGAGATATCTGATCGGAGCTTGGC AGGCGTTCAAGGCCCGCCCGACAATCCCACG
			10353386	CATCGATGGCGGCTGCAACGGCTGCGGTTCTGGT GGCCCTCGCGGCCGTAGTCCTCCTGTCCCT
QSB1_54x18_NR7	10359541	TGGTGGCCTGTAGGTTGTTGAGTTTGTCGTCAG GAGTCACTGGAAGGCCTTGAAGCTCGCCTGC	10277174	TTTCTAAACAATTTGAAGATGTTTGCAAACTTTGAT TACAGTTCGAAACCCACACATTATGTGT
			10339075	ATGATGCTAGGTCAGTGTACTCAGCTGCGAAGG TTGTCTATGATGGACCCATGTAAAATGCTCC

Rust_CT18_NR6	10573687	CACCAGCTGTGGCCGCACTTTCCACAGCTACGTT AGCAGCCGCTTCCACATCTTTAGCGACAAT	10325045	CGCAATGTTGGGGTCGAAACAGCTTGCGGAGCC CGAGTCTTGCTCGAGCGAGGCGTGGACGGGT
			10341636	GGGAGCTCCTGACCTACAATTGCATGCCTCTAGT TGTTTCTTCGGTGTTGGATTTTAAGGGCTT
Rust_CT18_NR10	10568868	AGTCTGTTGTCCGGAAAGCTGTGAACAAAGCGG TGGCCGGAATCCCGAGATCGGAAGAGCGGTT	10570318	TGGTCAATGAAGTAAATGAATTGGGAGAGACAG CTTTGTATACTGCTGCGGAGAAAGGGCACCT
			10326400	TTAAAAAGGTGCGAAGGTAATGCGGCTGTATCAA TCTCCTAGAAGCCTCTTAGGCTTCCGAGAT
Rust_CT18_HR3	10280956	AAAGCCAAGGAATTGAAAGTTTTGAACCTCAC CGAGATCGGAAGAGCGGTTCAGCAGGAATGCC	10274514	GAAGAAACGCAGGGAGAGAAGCTTCGGGCGTGC ATGGGGGGAAAGAAATCTGCCGAGATCGGAA
			10357059	GAAATGCTCCTGTTTGACCCGAGGAAGAGGATCA CGGTTACAGAGGCACTCCAGTGTCCTTACG
QSB1_CT18_HR3	10280176	CTCCCTACTGATTGATACTACAGGTAGCAATC ACTTCGCTCGCTGCACATCAGTTCTTTCTTCA	10285446	CTGGAAAAGGAGATGGCTTTCATCTTCAGATCCAC CGTGCATCTTCTTCGTTCGGCTAAAAAGG
			10279708	AGTCATCATCCACGTATTTTGACCTGCCTCGACAAC TTTACCCATTTCCCATCATCTTCCCGAA
QSB1_CT18_HR6	10305288	TGAGTATCTATAGAAGGTACTGGAATTTGCAT TTTTTTGGGAGCATACCGTGCTAGTGTCATTT	10280042	CCGATATTTGTCGAAATGGCAAGCCTGTCGACTTG AACCCTAGCAGGAGTTCCTGGAACTGGCC
			10276775	ATTTCGTCCATGTCTACGATACAAAAACTGGGTACA AGAAAGAGCAGGAAATTGACTTTTTCAG
QSB2_CT18_HR7	10575033	ACGACCGCCCGAAGACCACCGCTCCTTTCGCC GAACCGACGCCGACGCCATCGCAGGTACTCGC	10349358	AGAACGGCTCGATTCGTCCGTTTGAAGTATGTCCA CCGAGATCGGAAGAGCGGTTCAGCAGGAA
			End	

[a]Marker closest to the QTL peak. A second adjacent marker is shown when that QTL was unable to be placed by the marker closest to the peak location, so the closest flanking marker was used instead with a second to provide support for the positioning.

Supp. 4.2 Resistance QTL positions on individual *Corymbia* linkage maps



QTL with solid colour are significant, while those with hashed colour are suggestive.

Supp. 4.3 QTL for resistance to pathogens from *E. globulus* and *E. grandis* positioned in the *Corymbia* reference genome

QTL	Marker	Sequence	Chr	Position
Qmyco1	503461	ATTTCACACAGGAAACAGCTATGACCATGANNNNGCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTG GAATTCGCCCTTGATGGATCCAGTGCAGGACAGAAGGGCCAGAGAGCATCAAATAATCAGTACACATCCATACATAATTTAACACAT GGAAAGCTCACTCCGGAATTTGTATAAATCATTGTATAAATCAACTGTGCAATTATGCAATCTTCAATCAGCAGTGAAAGGATCCAAA ACATACAGTTTCCAAGACCACCCCATCCTTTACTAATGATGCCCCTTTGTCCAGTTGTCTCAAGAGCTTTGATTATGATGTCTGTCATTC TCTGTGGTTCTTGAACTGGCTGCTTATATCAACACAGAAGGAAAATCTATCAGAGCAGAAAAAGGAAACAAATCCAATGATCAGTGC ATAGATGCTGACAAGAAGGAACTCACAAGGCTACCAAAACCAATATAGATGGGCTTTTCACCTTCTTCTAGCCATTTCACTAGTGAAT CTGGTGGTTCATAACTTGATGCAAGGTCAAGAAAACAAAAGCCAACAACATCAATCTTGGGGCCCCAATCTGCACTGGATCCATCAA GGGCGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCAT	8	13,000,358
Qmyco2	565878	CAATTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTG GAATTCGCCCTTGATGGATCCAGTGCAGGAGAGTTTTTACGGACTAATTTGGATAATTTTTACTTCTTTTATGTCATTGATACTGTGTC GTGAATGTTCGTTTTGTTATCTAGAAACTGACTTGTGATTTTGGATAAAACCAGTTTCTCATGGTATGCGAGTCAAGGAACTTGCCTT GGCCATCTTAAACAGTTGGCAGCGCTTGATTATTAGGACAATGCTAACAAAGTGGCACCTCCCAGGAATTTGTCAATGTTTTGAGTA GCTCTCCTAGCTCGGAAGTGCTGGATATACACATCATCCCGGGTGAAGTGCCTTCCACTGATGTAAGTACACTTATCCCCATTTGAAG CTATTCTAGGTATGTGCGAAATTTCATGTAACTGTCTGCCTGATTTTGTAGACACAAGAAGCATGGCTGCGTTTGTGCAAAGATTCCT GAAGGGAGCTACTGGGATTTACGAGATCACTCTTTTGGATGTCTTACCAACTGCCTCAGAATTGTCAGAATGTCCGGCGACGTCAGG AAGGATCATGCGGTCCAGCTAGTGCAGTTTTTACTCTGTAAGACTTCCGTTTTACAGGAGCACGTCGTCTCTGCACTGGATCCATCAA GGGCGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCATGCATCTAGAGGGCCCAATTCGCCCTATAGTGAGTCGTATTA CAAT	1	23,056,741
Qmyco3	570894	ACGGCCGCCAGTGTGCTGGAATTCGCCCTTGATGGATCCAGTGCAGGTACGCTAGCAAATCTCCATTCATGACATTCTACTTTAGCTT GATAAATGTAAAAATGAGAACACCAAGAAGTGCTAGGACACCTTCTTTGAGGAGCCACTGCATCATGTATCTTGATACTTGTAAAAA AGTAGCAGTTAAATCACCTATATGAATATCATGATGGCGTTTCAAGGAATTATATGTATCACTAGCTCTTTCCTCTTACGCACCACAAT AAGTTGTGCGGCAAACACTACATGACCATGTTGAAGATAATTCGTAGTGGGCAATAGCATTTCCCTTTAGGTCCATATGCAACTATA GTGAACATCGTCCTAAATTCCTTGAGTGCTAAGATGGCGAGTCATGAATACACTTGCTATAGCGTGTATATTGAATTCACTGCACTGG ATCCATCAAGGGCGAATTCTGCAGATATCCATCACACTGG	7	10,300,487
Qmyco4	565163	CGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTGATGGATCCAGTGCAGACATCATAGCCGGTGATGGAGACTGG AGAGTGCTATTATGATACTTCACCCATCTGACATCTTCCTCTTCTCGCTTTCAGGCAATAAAATTGGGTCAATACTTCCTCTACGAACA ATTAACAAGTAGAGGTCTCTGGAACCGTAAAAGCAGAACTCAATCAAAGCTTATGAATATGTAAGTCTGTAGAACGTTTTTGTACGC CCTATACCCTTGCTCCATGACTGCTTCACAATCCCTTACTTAATGAAGTCTTCAGCCACTGTTACATTGAAAGAAAAGCAGAGCAAAAT	2	34,001,244

		ATGAGAGAGAGAGAGAGAGAGAGAGAGTTTCATAGACGAATGCAAACAACAAAGAAAGCATGAAAAAATGGAGAGGAAAAAAAT AGAGCTAGTAACCAGCTGACGATATAACATTGGGAAGTTGTTGATCCTTTTCTTGTCCTCAATCCATCCGATCAACAACCACCTTAACT GGTCTATAAGGCAAAGTACACCCTTCAGATTACTGTGGCCACTGCACTGGATCCATCAAGGGCGAATTCTGCAGATATCCATCACACT GGCGGCCGCTCGAGCATGCATC		
Qmyco5	644467	AAACTAATTTTTTCAACTCCGTCAAATGTTCATCATTTGATATATGGAGTGAAATACGAAGGTTAGTCACCAGCGCTTCCTGCTCCCAA CTCAAAGGTCCAGATGCGTGCAATGCCTCCATTGTGCAACGGTAAGCTTGCAACTCCAACTTGTGTATTTCATCTTCCAAGTTTTCTTT TGGCAAAGGGAGATTACTGCCTCCCTCATGTTCCCCATGACAAGTGGATTGAGCATCACCGATACAAGCGCTAACAAATTCAACAGG AGAAGCAGAGATGTCATGAGGACATATGTGACTATCATTTCCACTGATGCTACAACTACCGACTGAGCATTCTTCATTATTGTTGTGA CTGGATTCTGCACGGGATGCAAAAGCACAACCATCTGCATCGGTCAATTTAATCCTCTTCGCATCCACTTCAGTAAATCCATTTATTCT TCTCCTGAAAGAAGGCTTTAAGGTTTCCTGATCCAATGTGTCTCTTCGGATGGCAGCCGCATCAACTTGCTTCAATAGTGTAGGTGTC TTTGCAGCGGTTACATTGGGGAATCTTCCTTCTTTATCATTAACTTTACAACTCCTGAAAGCTTCAGAAGTTGCTTCTGGTACAATTGA GCCATTCAACCATGGTCTTTTCATAGATATATGTCTATTGACACAATAATCCTCCCAATTTTCGCCGTTCTCAGCTGTGGATAGATTATT TTTCTGAGACAGCTTCCGCGTTGCATGTTCCTTTTTACCTGGGAGGTTAGGATTCTTGTCACACCTCAACAATGGCATTTCCTTACATT TTCCTTCACGGAAAATTGCAGAATCCTGAATACAGAAACAATGTCAGAAAGTGAGACAAAAACAAATGCAAAGCAAGGGTTCACAT CTCTAATTCAAAACTGGTTTGCATTGTCCCACCAGTGATAAATAAACAAGTGTTCACATCTCTAACTGGTTTGCATTGTCCCATTTGTA GTAAACACGCAACAATATTAACATAATGACTAGCTTATGGCAGGATTCTAATTAA	3	8,729,047
Qmyco6	564786	GGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTGATGGATCCAGTGCAGTGGCCACAGTAATCTGAAGGGTGTACTT TGCCTTATAGACCAGTTAAGGTGGTTGTTGATCAGATGGATTGAGGACAAGAAAAGGATCAACAACTTCCTAATGTTATATCGTCAG CTGGTTACTAGCTCTATTTTTTTCCTCTCCAATTTTTCATGCTTTCTTTGTTGTTTGAATTCGTCTATGAAACTCTCTCTCTCTCTCTCT CTCTTTCTCTCTCTCCCTCCCTCTCTCTCTCATATTTTGCTCTGCTTTTCTTTCAATGTAACAGTGGCTGAAGACTTCATTAAGTAAGGGA TTGTGAAGCAGTCATGGAGCAAGAGTATAGGGCGTACAAAAGCGTTCTACAGATTTACATATTCATAAGCTTTGATTGAGTCCTGCT TTTACGCTTCCAGAGACCTCTACTTGTTAATTTGTTCGTAGAGGAAGTATTGACCCAATTTTATTGCCTGAAAGCGAGAAAAGGAAGA TGTCAGATGGGTGAAGTATCATAATAGCACTCTCCAGTCTCCGTCACCGGTCATGATGTCTGCACTGGATCCATCAAGGGCGAATTCT GCAGATATCCATCACACTGGCGGNCGCTCGAGCATGCATCT	2	34,001,538
Qmyco7	504015	CAATTTCACACAGGAAACANNNATGACCATGATTACGCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGC TGGAATTCGCCCTTGATGGATCCAGTGCAGTCTTAAGAATTTCACTCGTTCATGATGTCAACAGGTTATGAATGTTGATTCACTTAGT CCTCTCTTAAAGCAATCTATCATTAATTTAATCCTTACAACACCTGAAATTTCGTATCCCATTGAATTCGTCAAATATGAAGTTGACTTC GTGTGGTTCTGTAATTAAAGTAATTGCTTAGACAGGAAACACAAATTTTTTGGAGAGCGAACTGTTTCTTTCCAGCCATGAATTTTGA TCACAATAGACTGAAAAGGATTTTTTCCATCATTCTTTTCCGATATAAACAAATGCACTCAACATTTTCAGTCTGTAATTTCCTATGTCT TCGTGTTATCCTGTAGTACGTTCTATAATGATCGCTAGTCTGGAAACACAAGTATGATTGACTAAATTGTAACTCCCTTTCTTTGCAGC	11	18,823,562

		CACGAATCTTGATGGCAATGGCTAGGGATGGACTGTTGCCATCATTCTTTGGAGATATAAACAGACATACTCAAGTTCCGGTTAATA GTACGGTGATGACTGGTGTTCTGG		
Qmyco8	575308	GCTGATTTATGCTATTGTGGACTAAGTAGAATAAACTTTTGTAGAAAACTTGTCAGTATGTTGGAATCTTGATCTGTTGCCATTAAAT GAAACTGACCAGTGCTTCCAGATGTAAATCTGCCCATATAACTGCACTTGGTCTGGGATATGGCTACTAACCAGAATTTTAGGCTAA GTTTGATCGTGAAACTGATTTAGCTTAAAGCTTAGACTGTCCATTGTTAGAGTTAAACAAAAGCCAATGTGGCCAATTTGGTCTGACC CGCTAGGCATGTCAAATTTTTTCCTTTTAGCTTTTGTTTGATTTTCATCATCAGATTCATGAAAGCTAAATATATGGCTCATCAGATTCA TGAAAGCTAAATGGTAAGTTCTAATTCTTGTGGATATCTTTGCAGACATTGGAGGATGGAGATATGGCATTGTTTGATATGGGAGCT GAATATCATTTTTATGGATCTGATATTACCTGTTCATTCCCAGTGAGTTTCTGGTTCTTGTGTGCTTTTGTGGGGTTGACATGAGATAT TGTTTCTCCTAGAACTTCTTCCCACTGGATTACAATTTATTTTTTTGCAAATTTAATTTTGGTTTTCTGTTTTCCCTCAACTTATTTTTACA TTCTGCTATGGACTTACCCAATTTCTCTCAAAGGNNNAGNNNNTTTGATGATTTTTTTTTGGCCAGGTAAAAAAGCAAAAAAAAAAA AAAAAAAAAAAAATTCTCACGTATCTTCTTCATGGCTAAGGCCTGCTTATTCCTGAAGCTATTGTTATATTGATCCATGTCACTACAAT GGCATAGTTGAAAGGCTCTCCCAGGTTCAAACAAGTTGTTATTGGCTATTTAAAAAATTCATAATGTCTCAGTCATATTTTATGAATGT TTATCCTATATCTTCTGCCAGTGGGTGGTTTTTGTGGATAATTATTGATAGAGATGACAGAGATCTCCTTGGCTTTAGAAGGAAGCGG AGTTTACAAGATCAATTGTCTTCTTTTTGGATCAGAGTTACAAGTATCCCTTCTTACCATCAGCTTCATTTTGTGCAGGTGAATGGAAA GTTTACAAGTGATCAGAGTCTAATATACAATGTAAGTTGGTTACTTAAATTGACTATGAACAAAGCCAACTGTAGTTGCAGTAGAAAT TAGCTACGAACAAAGGCAACT	3	37,513,893
Sosr1	502976	AGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTGATGGATCCAGTGCAGTGAATTCAATATATACGCTATAGCAAGTGTATTCATG ACTTGCCATCTTAGCACTCAAGGAATTTAGGACGATGTTCACTATAGTTGCATATGGACCTAAAGGGAAATGCTATTGCCCACTACAA ATTATCTTCAACATGGTCATGTAGTGTTTGCCGCACAACTTATTGTGGTGCGTAAGAGGAAAGAGCTAGTGATACATATAATTCCTTG AAACGCCATCATGATATTCATATAGGTGGTTTAACTGCTACTTTTTTACAAGTATCAAGATACATGATGCAGTGGCTCCTCAAAGAAG GTGTCCTAGCACTTCTTGGTGTTCTCATTTTTACATTTATCAAGCTAAAGTAGAATGTCATGAATGGAGATTTGCTAGCGTACCTGCAC TGGATCCATCAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCAGATCGGAGATGGATCGAGAGGGCCGAGTTGACTTGATA ATGCGTGGTAATAGCCGTCACTGGCCTACGTTTTACAACGACATGACT	7	10,300,368
Sosr2	644396	GGCTATGTATTTTATCAATCGCTGCCCAGTAACAAGCTTACTGTCCCAACCATCCTCTTGTTTATTGGTGGAATCATCAAATACATAGA GCGAACATGTCCTCTGTATCTCGCAAGCTTTAGTAAATTCCGTCGCTCCTTGCTCAGCCTGCTGATGCCGGGCAGAACTATGCTAAAC TCATGGAAAAATACTCTTCCAAGAAGGAAGTTAATATTCCCGCTTCAATAGAGGTTATGCCGGAGCCCTACGTTCAATCCACAGATG GTGGAAAGACAGATGAAAAGATTTTGGACGACAGACAAGTGATGGAAGCCGCATTTTTTTACTTCACAACCTTCAAAGGCCTCCTTG TGGATCTAATCTTCAGGTTCTGTGAACGTGATGAAAGCATGAAATTTTTCAGAAGTAGAACCACGAAGGATGCTTTCCGGGTCATCA AGGTTGAGCTCAACTTCTTTTATGATATTCTTCTTACCAAGGCTTCCGGGGTTTACTATTTAACAGGGTGCCTTGTCCGAGCTCTCTCT CAATTGGTTCCGTCATTACTGCTTTTGCACTTTTCTATATACTGAACAAGCAGAGTTTTCGCGAATATGATACAAAGATTACATACACC TTGCTCCTTGGAGCTGTTGGCTTGGAATTTGTAGCTCTTAGTATGCTCATTTGCTCTAACTAGACAATTGCACTTTTAGGGAGATTGG	10	8,732,972

		AGAAACTTCGTAGATGTTCGCCTATCAAAAGCACCTTCATTGAGCTCCTCCTCAAATTCAAATGTGAAGATTCATCGTTTGCATTGCAT ATAGTCCATGCTAGATGGTCTAAATCTATCTTCTAGTACAATTTGATTGATTCTCGGCTGAGAAGGTGGCTAAAATGGATTGAGAAG CTCCTTGATCTCATCCCTAAGTGGCCAAAATGGATTGTGAAAATTCTTGATTTCATCTCTCTTAGGGAATTGTTTGATGACTTGAAGTG TGAACGAGAAGAACAGTACAGTGAAAAATTAGGTGAATTGATTTTTAAAGAGATGAAACGGAAGTCATCTGATGCAGAAGATTCAG AAAGCATGAAGAAAATGTGTGCTACTTGAGGTAGAAGGGCTCTGGAGAATACTAAAGCAAGGCAAGATTGCAAAGACTTGCTTCCA TTCATCTGTTATGTTGATTATGGTGAGAGTCTTTCACTGTGGCACATCACCACATAACTCTATTACAATGCTGATGCTAATACAAGTAC AAGCAGCGACCGTGAAAATAGTAAGATCTTGTCAGATTACATGCTTCCCTTATGATCAAACAACCTAATATGATGTC		
Sosr3	504787	GTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTGATGGATCCAGTGCAGGATACAAGGCAACTCAGAGCAAACATCATTATGCTACT AACACCCAACCAAAGGAAGATAATGTAAATGTCCTTGAACCCCAACCTTTGCATCTTGTAAGTACTCCAGAACCACACCAATTGTCTC ATTAAGTCCGTGGATCACCTTTGTGAAAGTATTCTCATCAATAAGTTCTCCTGTAAGATTGAAGAACAATGTCCATAAAAATCCCACA AGATCATGATATTAGACATGCATGATAACTATCGGTATCCATGACGTAGCTACTTACCTTCATCTGCACCCATATTTGCAACTAGTCTA ATTATCTTCTCAATCAAGGAAAAGGCAATAGCCAAAGTTCGCTGCTTTGAAAGGATAGAATCCACAGTGGAGGAGGAGCTCTTTGA GGCTTCATACTTCAAGTAGGCCATGTCATTCAGCAGTACAGCAATTTCAACCCTAGAAGATTCTAAGACAAGCAACAGACACCTAAA AGAGAATAGGGAACAATTTAGTGTAAGAACGAACCAAGATGCCAACATAACATCAAGTATTCCACTAAGAACCAATGATCAACTGA GAACAATAAGGCCAGTATGCATCACCTGTTGGCTGGAATGGGCTCCCTTGCATCAGGCGAGTTTACTTGATTAATGAGCCACTTCTC ACCCATTACTGAGATGGCAGACTCAGCCAAAATGAGAGCATGAAGCTTTTCAGCAGGAGCTAAATAGAAAAAGGAGCTACTTCAAC ATGCTTTCAATGAAAACCAAAAATTTTCCTCACTTAAGACACTCAAAAGGACATCATCAAAAGTTATTGTTAATTGGTAGTGAAAATA TCAGATATCAGATTAATTTTCCATCTAAACTAATGTTTAATTTAGCAATCTCTTGCCTTACAGGTTTGCTAAGAAAGTTACCTTGCTGCA CTGGATCCATCAAGGGCGAATTCTGCA	3	18,251,568
Sosr4	570163	AGCTCGGATNNCTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTGATGGATCCAGTGCAGTGGCTGTCTGATGTTGGTGTGAA ACATGTATGCCTGTATGATGAAGAAGGTGAAGAGGCTGTTTTATTTTCCATTTGATGTAAGTGCTGTACAAAAGTTTGTGATACTGTG GAAGTTACTCATGCTGTACCATTGACATCCTTTTGGGAGGGTTTAGGGGTCTTGAAGAAATCTAAGGATGCTATCTTGGAGAAGTTG AGCCATATCAGATTATTTAAGGTACTGAATATGGCATGTTCAACTTTGATACACTGCATTTGGTATTCTGCAATATGCTTGGTAATAG AGACTCATTCACATATATGTCGCTATGTAAACTGAACGTTATTGATATATTGGTTCAATTCATGAGGGTGTTTAAGACTGCCAAACGC TAGTGTTCTGCTGTTTGCTTTCTTTTGACTTTACATATTTTTGGGTAATTCAAAATGATAGTAATCATTCTTTAAATTATGAAATGTAGG TAACCTTTTATAGCTATTACCTTTTTATGATTTATCCAGCATGAAAATTTTAGTTCTTCTTAAAGATATCAGACTCTCCTATCTGCACTG GATCCATCAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCATGCATC	6	10,507,719
PPR2	572474	GAATTCGCCCTTGATGGATCCAGTGCAGCTCCTTATGGTAGGATTTCACAATTCCATAAGGTGATCCAAGATAGTTGCAGGTGTTGAT TGATCTGAGTAACTTTTCAATTGTGTGTCTGCTTGCATTCTTTCTCAACATGCTCACAACTATGCCACATGAAGAAGATGATTAGAACA GTAAATTGTGTGGAATACAGCGGTCGGTCAAGGAATTCTCACCTCAGAGGAAATGTATGCTGAGGAACGTCTTCTCAACAGGATGTT CCATCTGCACAGTGATTTTATTTTTAGCTCACACATCAACCCCCTCCAAATATGCTTCTTCATTGATGATTTTATTTTTGGCCTTTGGTCT	3	18,976,557

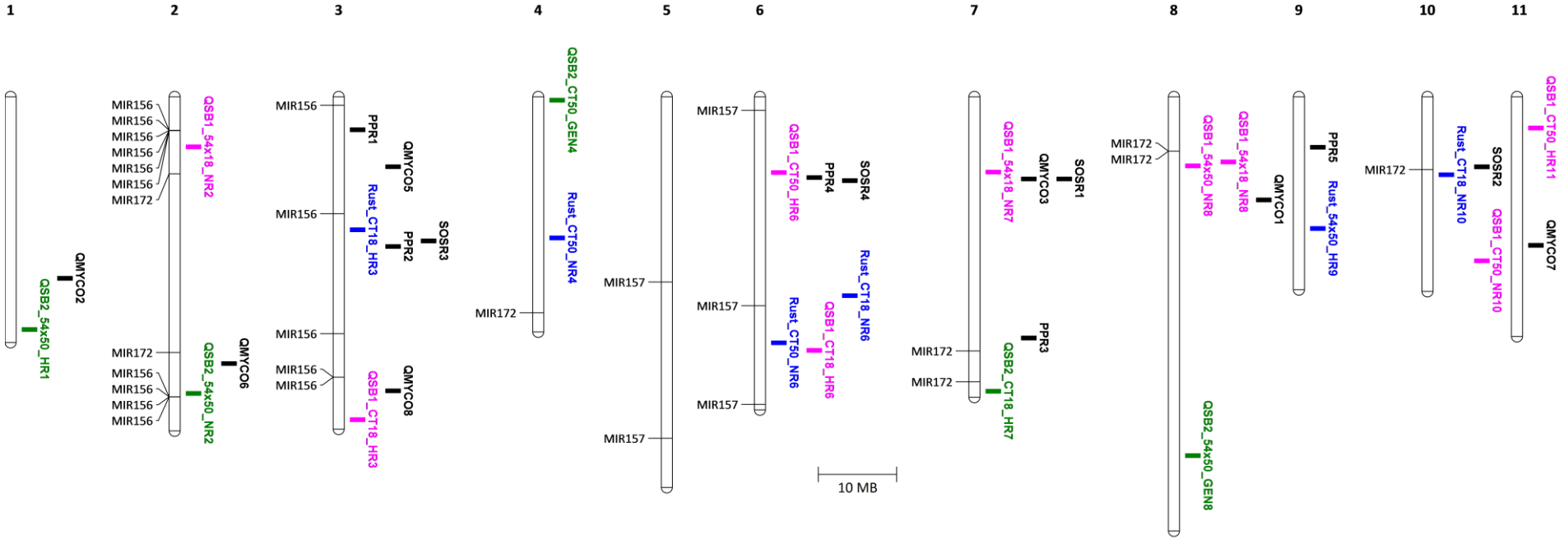
		ATGTCTAGTTGATCAAGAGGACCAAAGTATTATTAGGTATAACAGGAGTTAATAATTTGCTAACCAATGGAACAAATCTACCTTGCCT AGCATGAAATTGCCATGCTTGAATAATCATGACAAAAAAGAACATGACTGTCTGAAACAGATAAAACTGAATCAATATTCTTCAACT GGTCCAATTTGACACAGAGAAACACCACTAATGAATTCAAAGTTGCCTGGGAATAATATGAAGAGAACTTTGAGTTACACTTCTTTTA ATATCCTTTCTCAAACAAGTGTGCCAGAGAATTTTTAGTCTGCTGAAAGTACAAGGACAAGTTGAAGAATGGGATAAAGAGCTCTCA TGAGCTCTGCACTGGATCCATCAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGC		
PPR3	570240	CGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTGATGGATCCAGTGCAGGAAGGTCCAAGTGATCTGAT GCGACAAGCGCAGATCTTCTTATATGGAAATGGAAAGGATCAACCTCCTACTGCTATCAATCTGGAGAAAGCACTGTTGGAAGACTT GTTATTCAATCAAAGTCCTGCAAAGGTACATACTTAGAGAATGTAAATTACAGTGACGCAAAACAGCCCCACCACGACTCCTGAAAT GCTGTTAGACAGAAATCTGTTGCCTACTTCCTAAAAGCCCTTGAAGAGTTATTTGACGAATCTGGAATATGGATTTTCAGATAGCTTA ATACCTTCGTGAGTAGTATGACATTGGATAAGACTGCACTGGATCCATCAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCA GCTCGAGCATGCATCTAGAGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGA	7	30,718,368
PPR4	504554	GCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCC AGTGTGCTGGAATTCGCCCTTGATGGATCCAGTGCAGGTTCATCAGCCTCTCTTCCATTGCCTTTGATGGATCTGCCCAGTCGGTCTC TTCCTGCTTTAACCAACATGCCACGCCCCCTGAATGCAGAAGAGGGAGAGCCCATTACGGAGGTTCCACACGGTATGTGTGCATCTC TTAATTTTCCCTGAGTTAAAGCTCGTGCAGGTCGCTTGATTTGTCTTCTAGTCTTCACCTGAGGATAGGGATATAGGCAGAGAGAGA GAGAGAGAGAGGTGTCTATGTGTATTGCTATCATGTAAGTAAAGCTTTAATTCAAAGTATAATAACTGTGCTGCACTGGATCCATCA AGGGCGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCATGCATCTAGAGGGCCCAATTCGCCCTATAGTGAGTCGTATT ACAATTCACTGGCCGTCGTTTTACAACGTCGTGAC	6	10,118,232
PPR5	564752	GGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTGATGGATCCAGTGCAGAACTCTTTGACAAGCA TCAGGTGTTAAGCTATAAAAAGCACACCTGGGAAGAAGAAGATACATCCATTGTGTGCTCCTTTGAAAGAAAAGGACATGAGGACC CTGAATTACACTAAACCACAATAACAAGAAGCAAAGCTGACATGTTTAACCTCTGGCCACATGCCACAATACACTAGCCATAATTAAT TCCACATGTACATACCCTTTGTACTTAGAAAGGCAAACAATCTGTTGGACGGGAACAACATCCAAGTATGCAGACAGAAAGGCATCA AGAGAAAAAGGACAAAACATAATTGCCAAACCGGCAAACATATACCTACATCTCTATGCTCTGAGATGAATTTGTCCAAGTAATTGA TCAACGCAATCTGCTCAGTATAACTGCACTGGATCCATCAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCAT	9	6,224,560
PPR1	Emb1071 forward primer	CAACTGTTATTGAGAATGTATCGAA	8	6,343,438
PPR1	Emb1071 reverse primer	AACCTTACCTGCACCCTCCT	8	6,343,171

These QTL are taken from Junghans et al. (2003a); Freeman et al. (2008b); Mamani et al. (2010); Butler et al. (2016).

Supp. 4.4 Putative annotations of gene models underlying QTL for resistance in *Corymbia*

This material is unfeasible to present in this thesis due to its format. It can be found at:
http://ecite.utas.edu.au/127935

Supp. 4.5 Co-location of microRNAs with QTL for resistance in *Corymbia*



Chapter 5 - Annotation of the *Corymbia* terpene synthase gene family shows broad conservation but dynamic evolution of physical clusters relative to *Eucalyptus*

Introduction

Terpenes are an extensive group of hydrocarbon-based compounds present in most plants, with thousands currently characterised (Padovan et al. 2014). While some terpenes are present in essentially all plants as primary metabolites, such as gibberellin or abscisic acid (Chen et al. 2011), many are secondary metabolites. Correspondingly, there is wide variation in the terpenes produced across different species, in line with their role in modulating diverse interactions between plants and their environment (Keszei et al. 2010a). Along with regulating growth and other developmental processes (Chen et al. 2011), terpenes play roles in pollinator attraction (Pichersky and Gershenzon 2002), chemical and physical barriers to herbivory (Lawler et al. 1999; O'Reilly-Wapstra et al. 2004; Heiling et al. 2010), and thermotolerance (Peñuelas et al. 2005), to name a few. Terpenes are also important economically due to their utilization as pharmaceuticals, industrial materials and biofuel precursors, as well as their direct impact on the fragrance and flavour of horticultural food products such as apples and wine (Schwab et al. 2013).

These varied terpenoid products are created by terpene synthase (TPS) enzymes. TPS enzymes synthesize terpenoid products from isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which are both created by the action of the mevalonic acid (MEV) pathway operating in the cytosol and the methylerythritol phosphate (MEP) pathway operating in the plastids (Chen et al. 2011). Extensive study in many plant species (Aubourg et al. 2002; Martin et al. 2010; Xiong et al. 2016; Hansen et al. 2017) has revealed that the *TPS* gene family is generally a mid-size family, with gene numbers ranging from 1 in *Physcomitrella patens* to 113 in *Eucalyptus grandis*. Previous phylogenetic analyses of the *TPS* gene family have revealed eight different subfamilies, designated *TPS-a* through to *TPS-h*. Each subfamily influences the synthesis of different terpenoid products, with the genes in the subfamilies *TPS-a* (sesqui-terpene), *TPS-b* and *TPS-g* (cyclic/acyclic mono-terpene), *TPS-c* and *TPS-e* (copalyl diphosphate, ent-kaurene, and di-, mono- and sesqui-terpene) and *TPS-f* (ent-kaurene and di-, mono- and sesqui-terpene) categorized by the structurally distinct compounds they synthesize. Subfamilies *TPS-c*, *-e* and *-f* are predominantly involved in the synthesis of primary metabolites such as gibberellin and abscisic acid, while

subfamilies *TPS-a, TPS-b* and *TPS-g* generally synthesize secondary metabolites including cineole and citronellal (Chen et al. 2011). The representation of *TPS* subfamilies is quite different across taxa; *TPS-d* and *TPS-h* subfamilies, for example, are specific to gymnosperms and *Selaginella* spp., respectively (Chen et al. 2011). Given the large amount of variation in terpenoid profiles within and between taxa as well as the economic and evolutionary importance of terpenoid products (Keszei et al. 2008; Schwab et al. 2013), it is no surprise there is an extensive body of literature exploring these compounds. Further investigation of the gene family underlying terpenes in different taxa will greatly contribute to understanding how this diversity arises.

The Myrtaceae, of the order Myrtales, are a family of plants that exhibit some of the highest concentrations and diversity of foliar terpenes in plants. Across the Myrtaceae, hundreds of compounds have been characterised (Padovan et al. 2014), with the foliage of individual trees often containing over 40 identifiable compounds (Keszei et al. 2008). Due to these features many Myrtaceous genera are key resources for commercial industries utilizing terpenes as essential oils (Padovan et al. 2014) including *Melaleuca* (Keszei et al. 2010b), *Leptospermum* (Douglas et al. 2004), *Eucalyptus* and *Corymbia* (Batish et al. 2008). Along with *Angophora*, *Eucalyptus* and *Corymbia* are broadly classified as eucalypts (Slee et al. 2006). Eucalypts are the dominant trees in most Australian native forest and the predominant hardwood plantation species in Australia and overseas, due to their importance to the pulp, biofuel and timber industries (Rockwood et al. 2008; Shepherd et al. 2011). The characteristic smell of the eucalypts is due to their especially high concentration of foliar terpenes, with the high diversity of compounds present in this foliage extensively studied (Ammon et al. 1985; Lawler et al. 1998; Asante et al. 2001; Keszei et al. 2008). While the terpenoid profiles of most eucalypts are dominated by α-pinene and 1,8-cineole (Keszei et al. 2010a; Padovan et al. 2014), chemotype variation is important both to plant ecology (O'Reilly-Wapstra et al. 2004; Keszei et al. 2010a) and the essential oil industry. Analysis of the *Eucalyptus grandis* reference genome (Myburg et al. 2014) revealed that this variability is accompanied by the largest number of *TPS* genes of any plant yet sequenced, closely followed by *Eucalyptus globulus* (Külheim et al. 2015). These genes often occur in duplicate arrays or physical clusters which are prone to relatively rapid expansion and contraction (Hanada et al. 2008). Given the most likely fate of duplicate genes is degeneration (Lynch and Conery 2000), the large number of genes present in these eucalypts suggests natural selection preserved these expansions, resulting in high variability of terpene products. Indeed, the combinations of terpenes present in eucalypts varies both between and within species (Keszei et al. 2008; O'Reilly-Wapstra et al. 2011) and within individuals (Padovan et al. 2012), in line with the diverse roles these

compounds play in responding to ecological variation. Specific comparison of *E. grandis* and *E. globulus* also suggests that most *TPS* genes evolved prior to the divergence of these species, approximately 12 million years ago (MYA), but points to ongoing evolution as indicated by novel gene duplication, degeneration and gene loss (Külheim et al. 2015). Although this gene family has been well categorized in *E. globulus* and *E. grandis*, the extent to which it is conserved in other eucalypt lineages is currently unknown.

The genus *Corymbia* is predominantly endemic to the tropical, arid, and semi-arid zones of northern Australia (Hill and Johnson 1995; Ladiges et al. 2003), but is increasingly cultivated for forestry and essential oil production in Australia, India, Brazil, Fiji and South Africa (Asante et al. 2001; Vernin et al. 2004). It is a sister genus to *Eucalyptus* (Lee 2007), which diverged from a common ancestor approximately 52 MYA (Crisp et al. 2011; Thornhill et al. 2015). All eucalypts share the same haploid chromosome number (n = 11), which is highly conserved across most Myrtaceous species (Grattapaglia et al. 2012). However, in comparison to *E. grandis, Corymbia citriodora* subsp. *variegata* (hereafter referred to as CCV) has both a smaller genome size (370 MB vs 640 MB, Grattapaglia and Bradshaw Jr 1994) and several major differences in chromosome structure (Butler et al. 2017b). The recent *de novo* genome assemblies for two CCV genotypes (Shepherd et al. 2015) provides the opportunity for comparison of individual loci and gene families.

In this study we annotate the *terpene synthase* gene family in the CCV reference genome and compare it to other plants (*Vitis, Populus, Arabidopsis*), but focus on the comparison of *Corymbia* with *E. grandis* and *E. globulus* (Külheim et al. 2015). We present evidence for broad conservation in this gene family across eucalypt lineages along with extensive variation within subfamilies in terms of the presence of specific clusters, and the number of genes contained within them. These results are discussed in the context of their evolutionary and ecological importance.

Material and methods

Terpene synthase gene discovery

Initially, a CoGeBLAST (Lyons et al. 2008) search for *TPS* genes was performed on the CCV reference genome v1.1 (CCV18, Healey et al. 2017), based on conserved domains from all *TPS* subfamilies following Külheim et al. (2015). A preliminary list of putative *TPS* genes was created based on hits with high similarity (e-value < $1e^{-08}$). To identify if these preliminary hits were full length genes, the genomic regions surrounding each BLAST hit (± 5,000 bp) were used in reverse BLAST searches against the non-redundant database at Genbank (http://www.ncbi.nlm.nih.gov, accessed 23/02/2017). The closest matching *TPS* gene from

E. grandis, E. globulus, Arabidopsis thaliana, Populus trichocarpa or *Vitis vinifera* was compared to the putative *TPS* sequence using GeneWise (Birney and Durbin 2000), to determine exon-intron borders and reveal reading frame shifts or premature stop codons. A partial genome assembly from a different CCV individual (1CCV2-054) was also mined for *TPS* genes and, where possible, used to validate the results from the CCV18 genome assembly (Healey et al. 2017).

Phylogenetic analysis and annotation

The amino acid sequence of all putative CCV *TPS* genes were aligned using ClustalW along with those from *E. grandis, E. globulus, A. thaliana, P. trichocarpa* and *V. vinifera* (Külheim et al. 2015). Due to high levels of variation and variable exon counts between taxa the alignment was trimmed to focus on regions conserved among all genes (positions in the alignment with > 75% gap representation were removed), allowing a direct comparison with the results of Külheim et al. (2015). The phylogeny of the *TPS* family in these six organisms was determined using IQTREE (Nguyen et al. 2015) with 1,000 ultrafast bootstrap replicates (Minh et al. 2013). The JTT amino acid substitution model with estimation of invariable sites and gamma distribution was used as this model created the tree with the highest AICc value (corrected Akaike's information criterion) using the program SMS (Lefort et al. 2017). CCV putative *TPS* genes were sorted into the subfamilies *TPS-a, -b, -c, -e, -f* and *-g* based on sequence similarity to *TPS* genes previously classified in the other species. These genes were sorted by chromosome and by position within chromosome in the CCV reference genome, and annotated from the first *TPS-a* gene (*CorciTPS001*) to the final *TPS-g* gene (*CorciTPS102*).

Gene birth/death rates were estimated using the program Badirate (Librado et al. 2012). The BD-FR-CML model was used with the family option, which allows for a free turnover rate for each branch of the species tree, with the gain/loss events of internal nodes inferred by maximum likelihood and informed by the relative representation of each subfamily (Librado et al. 2012). This was performed for both the tree of the six species (with divergence times taken from Wikström et al. (2001)), and for the eucalypts alone. To improve the accuracy of the rate estimation in the latter analysis of the eucalypts, *TPS* subfamilies were further divided into their component orthologous groups before analysis, which were defined as the most inclusive clade of the gene tree compatible with the species tree.

RNA-Seq expression analysis

To examine the expression of putative functional genes, RNA sequencing was undertaken using mRNA isolated from five tissue types: flower initials, flower buds, bark, expanded leaf, and unexpanded leaf. Tissue was obtained from 1CCV2-054 (sequenced for the CCV54 assembly), and RNA extracted using Ambion RNAquenous kit with Ambion RNA Isolation aid and the standard protocol (Life Technologies Australia, Mulgrave Vic). Total RNA was shipped to AGRF (Melbourne, Australia) for library preparation (TruSeq Stranded mRNA Sample, Illumina) and sequencing (HiSeq HT chemistry single read 50/100, Illumina). A total of 75 GB of sequence data was generated across all five libraries: 25 GB of 100 bp single-end reads, and 50 GB of 100 bp paired-end reads. Reads were quality controlled using BBMap tools (Bushnell 2016), and assembled into transcripts using Trinity *de novo* RNA-Seq assembly pipeline (Haas et al. 2013). Transcripts were aligned to the CCV reference genome using CoGe's RNA-Seq analysis pipeline (Lyons and Freeling 2008). Detectable expression at the location of putative functional and pseudogenes was a criteria used to support the existence of putative genes. The clustering of gene expression was examined using the complete linkage method and Euclidean distance measures contained within the package 'gplots' (Warnes et al. 2016) in R (R Core Team 2017), allowing clusters to be identified based on dendogram structure.

Comparative analysis of the TPS gene family between species

To examine differences in genome organization and gene number in specific *TPS* clusters, the positions of *TPS* genes in the CCV and *E. grandis* genomes were collated and assigned to specific physical clusters. A physical cluster of *TPS* genes was defined as genes from the same subfamily occurring on the same chromosome, with further support for gene clusters based on close phylogenetic relationships. Homologous clusters were matched, requiring both close phylogenetic relationships between *TPS* genes and similar genomic position in each genome assembly. Homologous clusters that were both syntenic (located on the same chromosome) and matched the approximate position within that chromosome in both species were examined for copy number variation. *TPS* genes in the CCV54 assembly were also assigned to physical clusters and compared to the CCV18 reference genome to determine if there were any changes in copy number.

In cases where gene clusters in the CCV reference genome were placed on a different chromosome to their apparent homolog in *E. grandis* (evaluated by phylogenetic relatedness), verification of their position was undertaken in CCV54. The tool 'SYNFIND' in CoGe (Lyons and Freeling 2008) was used to determine the likely position of homologous

genes in CCV54 taking into account the synteny of the surrounding region. If gene position was conserved across both CCV genome assemblies, movement of loci relative to *E grandis* was considered real, while disagreements between the CCV assemblies were flagged as possible errors caused by misassembly, with more weight given to the loci position mirroring that of *E. grandis*.

As the CCV genome assemblies (Healey et al. 2017) are anchored to linkage maps (Butler et al. 2017b), it is possible that markers on these maps may be mis-ordered, leading to incorrect contig positioning and potentially incorrect conclusions on loci position and movement. To examine this, the number of markers used to anchor and orient each contig housing *TPS* loci with putative movement was used to determine the strength of contig placement.

Results

Discovery of TPS loci

In the *Corymbia citriodora* subsp. *variegata* reference genome (CCV18) 127 loci were discovered with high sequence similarity to terpene synthase (*TPS*) genes from other species. Using a modified version of the classification method of Külheim et al. (2015), loci were classified into three categories: (i) 64 were full length with no structural abnormalities and had evidence of expression; (ii) 17 were full length, expressed but with up to two frame shifts or premature stop codons; and (iii) 21 were full length, had no evidence of expression and up to two frame shifts or premature stop codons. In accordance with Külheim et al. (2015), these were considered putatively functional *TPS* genes, resulting in a total of 102 genes (Table 5.1, Supp. 5.1) used in further analysis. The remaining 25 loci were classified as pseudogenes with more than two frame shifts or premature stop codons, with no consideration given to expression (Supp. 5.2). Similar analysis of the partially assembled CCV54 without expression data revealed 64 putative functional *TPS* genes and seven pseudogenes (Supp. 5.3).

Table 5.1 Copy numbers of *TPS* genes by subfamily in various plant species

Subfamily	*C. citriodora* subsp. *variegata* (CCV)	*E. grandis*	*E. globulus*	*V. vinifera*	*A. thaliana*	*P. trichocarpa*
TPS-a	51 (2)	52	45	29	23	13
TPS-b1	26 (1)	27	28	8	6	10
TPS-b2	10 (1)	9	10	2	0	2
TPS-c	1 (1)	2	2	2	1	2
TPS-e	1 (1)	3	2	1	1	2
TPS-f	4 (0)	7	9	0	1	1
TPS-g	9 (0)	13	10	15	1	2
Total	102	113	106	57	33	32

Table adapted from (Külheim et al. 2015). Numbers in brackets indicate the number of orthologous pairs between *C. citriodora* subsp. *variegata* (CCV) and *E. grandis.* See Figure 5.1 for examples of orthologous pairs.

Phylogenetic analysis

The phylogenies presented show the relationship between the CCV18 *TPS* genes and those from *E. globulus, E. grandis, V. vinifera, P. trichocarpa,* and *A. thaliana*, divided into *TPS-a* (Figure 5.1), *TPS-b* and *TPS-g* (Figure 5.2) and *TPS-c, TPS-e* and *TPS-f* (Figure 5.3) subfamilies. The same *TPS* subfamilies were represented in each eucalypt species. Orthology (genes in different species directly descended from the same ancestral gene) between *TPS* genes in *E grandis* and *E. globulus* was common, with 60% of genes found in orthologous pairs (defined as a single gene in one species more closely related to a single gene in a different species than to a gene within its own genome, see Figure 5.1 for examples). However, only 9% of *TPS* genes in CCV were orthologous with pairs from the other eucalypts.

The *TPS-a* subfamily was represented by the most genes in CCV, as was the case in *E. grandis* and *E. globulus* (Table 5.1). However, specific *TPS-a* clades in CCV were expanded relative to the other eucalypts (for example, the clade containing *CorciTPS035* [Figure 5.1-a]), or missing entirely (for example, the clade containing *EgranTPS029* [Figure 5.1-b]). An interesting orthologous relationship was seen between an *E. globulus TPS* gene (*EglobTPS022*) and a clade of CCV *TPS* genes with no specific *E. grandis* ortholog, suggesting this gene was lost or not found in *E. grandis* (Figure 5.1-c). While 31 of the *TPS-a* genes in *E. grandis* (60% of total *TPS-a* genes) and *E. globulus* (69%) were in orthologous pairs, greater divergence was evident in CCV as only two *TPS-a* genes (4%) were in orthologous pairs with other eucalypt genes (specifically *CorciTPS025* and *CorciTPS026*; Figure 5.1).

As seen in the *TPS-a* subfamily, the *TPS-b* and *TPS-g* subfamilies also provided evidence for expansion and contraction of physical clusters as well as loss of loci among the eucalypts (Figure 5.2). Only one *TPS-b1* gene (*CorciTPS053*) in CCV (4% of the total) was in an orthologous pair with the other eucalypts. In contrast, 19 of the *TPS-b1* genes in *E. grandis* (70%) and *E. globulus* (68%) occurred in orthologous pairs. Another potential gene loss in *E. grandis* was seen in the clade containing *EglobTPS077* and multiple CCV genes (Figure 5.2-a). Of the *TPS-b2* genes (Figure 5.2), five were in orthologous pairs between *E. grandis* (55%) and *E. globulus* (50%), while in CCV only one (10%) was orthologous to the other eucalypts. The remainder of the genes were arranged in clades specific to each eucalypt with no orthologous pairing (Figure 5.2-b, 5.2-c). In the *TPS-g* subfamily, six genes in *E. grandis* (46%) and *E. globulus* (60%) were found in orthologous pairs, but no orthologous pairs were found between CCV and the other eucalypts.

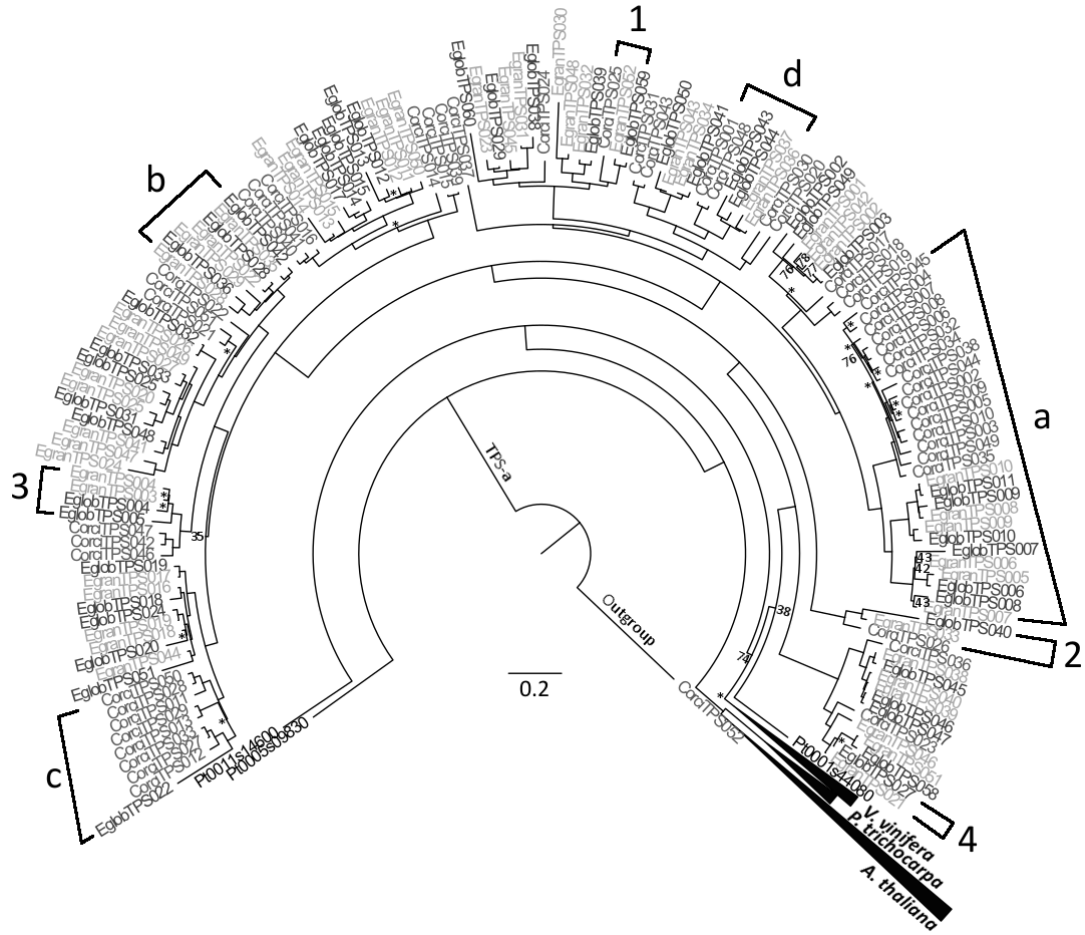


Fig. 5.1 Phylogeny of the *TPS-a* subfamily. This tree was created through maximum likelihood analysis comparing the *TPS-a* subfamily from *C. citriodora* subsp. *variegata* (Corci) with those from *E. grandis* (Egran)*, E. globulus* (Eglob)*, P. trichocarpa* (Pt), *V. vinifera* (Vv) and *A. thaliana* (At). Bootstrap values supported by < 80% are noted by number, while those with bootstrap values between 80 - 94% are indicated by the symbol *. All others have values > 95%. Scale represents amino acid substitutions per site. A *TPS-b* gene from *C. citriodora* subsp. *variegata* was used as the outgroup. a - d refers to results discussed in the text. Examples of orthologous pairings are given by numbers 1 & 2. 3 is not considered an orthologous pairing as *EglobTPS004* shares its most recent ancestral gene with two genes from *E. grandis* rather than one. 4 is not considered an orthologous pairing as *EglobTPS027* and *EgranTPS021* do not share the same most recent ancestral gene. b gives an example of genes in orthologous pairings, with the exception of *EgranTPS029*, which is does not pair to a single gene from another species. c shows an example of a non-orthologous pairing, as *EglobTPS022* is closely related to several genes from CCV rather than a specific one.

Fig. 5.2 Phylogeny of the *TPS-b* and *TPS-g* subfamilies. This tree was created through maximum likelihood analysis comparing the *TPS-b* and *TPS-g* subfamilies from *C. citriodora* subsp. *variegata* (Corci) with those from *E. grandis* (Egran)*, E. globulus* (Eglob)*, P. trichocarpa* (Pt), *V. vinifera* (Vv) and *A. thaliana* (At). Bootstrap values supported by < 80% are noted by number, while those with bootstrap values between 80 - 94% are indicated by the symbol *. All others have values > 95%. Scale represents amino acid substitutions per site. A *TPS-a* gene from *C. citriodora* subsp. *variegata* was used as the outgroup. a - c refers to results discussed in the text.

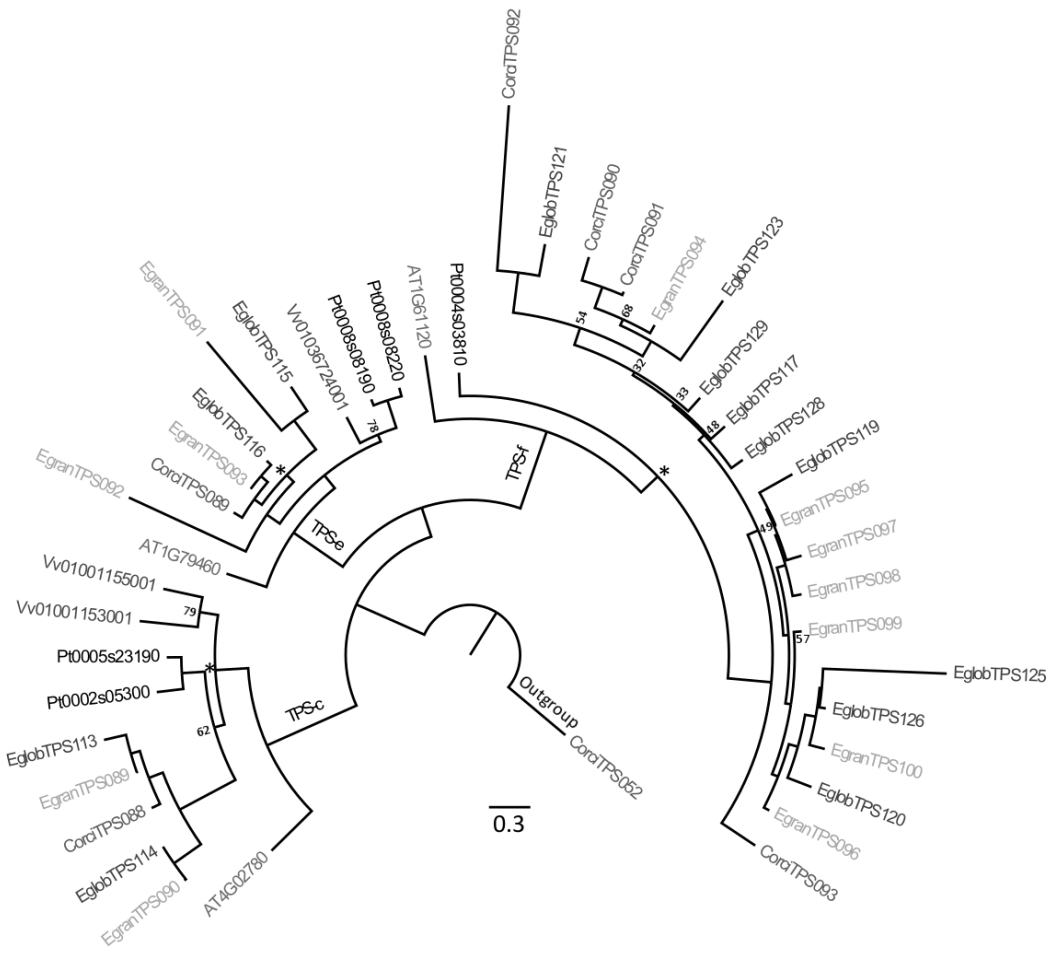


Fig. 5.3 Phylogeny of the *TPS-c, TPS-e* and *TPS-f* subfamilies. This tree was created through maximum likelihood analysis comparing the *TPS-c*, *TPS-e* and *TPS-f* subfamilies from *C. citriodora* subsp. *variegata* (Corci) with those from *E. grandis* (Egran)*, E. globulus* (Eglob)*, P. trichocarpa* (Pt), *V. vinifera* (Vv) and *A. thaliana* (At). Bootstrap values supported by < 80% are noted by number, while those with bootstrap values between 80 - 94% are indicated by the symbol *. All others have values > 95%. Scale represents amino acid substitutions per site. A *TPS-b* gene from *C. citriodora* subsp. *variegata* was used as the outgroup.

The *TPS-c* and *TPS-e* subfamilies, involved in the synthesis of primary metabolites, were generally conserved between the eucalypts (Figure 5.3). The single *TPS-c* gene in CCV was found in an orthologous pair with both other eucalypts, while a second orthologous pair was found between *E. grandis* and *E. globulus*. An identical situation was observed in the *TPS-e* subfamily, with the single gene in CCV paired with the two *Eucalyptus* species, and a second orthologous pair between *E. grandis* and *E. globulus.* In both cases, a second *TPS-c* and *TPS-e* gene was found in the CCV54 assembly in the minor scaffolds (contigs that were assembled into scaffolds but not anchored to the 11 chromosomes), suggesting the corresponding genes may be missing from the CCV18 assembly (although the possibility that the minor scaffolds represent alternate haplotypes which did not fuse to the chromosomes cannot be dismissed). Both of these subfamilies are highly conserved in *A. thaliana, V. vinifera* and *P. trichocarpa*, as each only has 1 - 2 genes of each subfamily (Figure 5.3).

The *TPS-f* subfamily was more dynamic than the other subfamilies involved in primary metabolism (Figure 5.3). The orthologous pairings seen in this clade differed somewhat to those presented by Külheim et al. (2015), likely influenced by low bootstrap support in both studies, slight differences in methodology and the addition of CCV weakening support for previous clade structure. In our analysis, only two of the *E. grandis* (29%) and *E. globulus* (22%) *TPS-f* loci were in orthologous pairs, while a single *TPS-f* loci was directly orthologous between CCV (25%) and *E. globulus* (*CorciTPS092* and *EglobTPS121*), with no gene from *E. grandis* present. In contrast to *TPS-c* and *TPS-e*, *A. thaliana* and *P. trichocarpa* only have a single *TPS-f* gene, while no *TPS-f* was found in *V. vinifera* (Table 5.1).

The estimated gene birth rate in the *TPS* gene family was negligible ($\leqslant$ 0.0002 events/gene/million years [e/g/my]) for *A. thaliana, V. vinifera* and *P. trichocarpa*, while the death rate ranged from 0.0016 - 0.0031 e/g/my (Supp.5.4-a). In contrast, the eucalypt lineage was estimated to have experienced a magnitude higher rate of gene birth (0.0282 e/g/my). Within the eucalypt lineages, death rate was similar in both *Eucalyptus* and *Corymbia* (0.0063 - 0.0071 e/g/my, Supp. 5.4-b). However, the gene birth rate in *E. grandis* (0.0125 e/g/my, since divergence from *E. globulus*) was seven times higher than the estimated birth rate in CCV (0.0018 e/g/my).

Proportional representation and genome organisation of TPS genes

There were no significant differences in subfamily representation (the proportion of genes in each subfamily) between *E. grandis* and CCV ($\chi^2_4 = 3.69$, $P > 0.05$ [combining *TPS-c*, *-e*, and *-f* due to sample size]), or the number of genes involved in primary *versus* secondary

metabolism ($\chi^2_1 = 2.41$, $P > 0.05$). A similar lack of significant difference was observed between *E. grandis* and *E. globulus* in the number of loci at the subfamily ($\chi^2_4 = 1.53$, $P > 0.05$) or primary *versus* secondary metabolite ($\chi^2_1 = 0.3$, $P > 0.05$) levels, providing evidence that the broad features of this gene family are conserved between *Eucalyptus* and *Corymbia*.

Seventy-five putative functional *TPS* genes were found across all 11 chromosomes (74% of the total) in the CCV18 genome assembly, with 27 genes found within minor scaffolds (26%, Supp. 5.1). In comparison, 97 and 16 *TPS* genes were found on chromosomes (86%) and minor scaffolds (14%), respectively, in *E. grandis.* The relative proportion of genes located on the main chromosomes in each species is consistent with the estimated completeness of each assembly (Myburg et al. 2014; Healey et al. 2017). In the CCV genome, *TPS* genes were often arranged in physical clusters with genes from only one subfamily, as was seen in *E. grandis* (Külheim et al. 2015). On average, there were 3.7 *TPS* genes per cluster in CCV (only considering those on chromosomes), while *E. grandis* averaged 5.1 per cluster. This difference may reflect the greater proportion of *TPS* genes on minor scaffolds in CCV compared to *E. grandis*. In CCV these clusters occurred in true tandem arrays (no intervening genes between putative *TPS* genes), localised clusters with other genes contained within and combinations of the two (Table 5.2, Supp. 5.5).

Table 5.2 Structure of the *Corymbia citriodora* subsp. *variegata* (CCV) terpene synthase physical clusters and the *Eucalyptus grandis* clusters which are syntenic to CCV

Species[a]	Subfamily	Chr	*TPS* genes	Position (bp)	Span (bp)	Intervening genes	Internal clustering[b]
CCV	*TPS-a*	Chr3	9	29,395,680	3,598,941	202	1(9)2(1)1(22)3(2)1(168)1
Egr	*TPS-a*	Chr3	3	47,928,331	1,606,854	64	1(2)1(62)1
CCV	*TPS-a*	Chr4	3	11,823,925	14,705	0	N/A
CCV	*TPS-a*	Chr4	2	18,100,282	23,876	0	N/A
Egr	*TPS-a*	Chr4	5	19,896,024	246,197	7	1(2)3(5)1
CCV	*TPS-a*	Chr5	4	16,525,672	1,219,850	68	1(68)3
CCV	*TPS-a*	Chr6	6	33,461,887	106,929	0	N/A
Egr	*TPS-a*	Chr6	10	42,991,263	321,452	0	N/A
CCV	*TPS-a*	Chr7	3	2,183,056	809,251	59	1(58)1(1)1
CCV	*TPS-a*	Chr7	2	14,463,251	1,761,379	105	1(105)1
CCV	*TPS-b1*	Chr1	8	22,285,470	1,866,767	103	1(64)2(1)2(36)1(1)1(1)1
Egr	*TPS-b1*	Chr1	7	17,720,921	1,286,126	50	1(3)4(1)1(46)1
CCV	*TPS-b1*	Chr2	2	2,027,307	13,692	0	N/A
CCV	*TPS-b1*	Chr4	2	8,861,360	169,636	5	1(5)1
Egr	*TPS-b1*	Chr4	8	16,009,931	217,347	7	1(1)1(1)2(4)2(1)2
CCV	*TPS-b1*	Chr5	3	41,952,144	3,787,929	215	1(192)1(23)1
CCV	*TPS-b1*	Chr8	3	29,799,505	47,398	3	1(1)1(2)1
CCV	*TPS-b2*	Chr10	3	13,605,402	38,144	1	2(1)1
CCV	*TPS-b2*	Chr11	5	23,810,145	420,788	25	2(25)3
Egr	*TPS-b2*	Chr11	9	10,288,308	1,164,219	33	3(31)1(1)4(1)1
CCV	*TPS-f*	Chr4	3	7,135,889	385,574	31	1(31)2
Egr	*TPS-f*	Chr4	7	12,270,273	287,241	2	3(2)4
CCV	*TPS-g*	Chr2	2	19,657,471	29,412	2	1(2)1
CCV	*TPS-g*	Chr5	3	44,341,209	100,371	2	1(2)2
Egr	*TPS-g*	Chr5	12	62,540,499	1,272,677	25	3(2)3(2)1(19)1(1)1(1)3

[a]Egr indicate *TPS* clusters from *E. grandis* syntenic with the CCV *TPS* cluster directly above. Loci with only a single *TPS* gene are not shown.
[b]Structure of the gene cluster, with non-*TPS* genes indicated by brackets. N/A indicates no intervening genes.

Across all *TPS* subfamilies, 10 physical clusters were both syntenic and phylogenetically similar between *E. grandis* and CCV18 (Figure 5.4). These clusters were assumed to be homologous between these species and were examined for copy number variation. *E globulus* was not examined due to the lack of an assembled genome. There was no significant correlation between gene number in syntenic homologous clusters between species (Spearman's $r_8 = 0.29$, $P > 0.10$), suggesting independent expansion or contraction has occurred between *E. grandis* and CCV. Seven clusters were homologous but non-syntenic, with the chromosome assignment of three non-syntenic clusters in the CCV

reference genome supported by the second CCV genome assembly CCV54 (Figure 5.4, Supp. 5.6). The position of the single *TPC-c* gene conflicted between the CCV assemblies (despite both CCV18 and CCV54 [not shown] having contig-marker support for placement), potentially due to assembly error in one or the other. The general placement of clusters was supported by examining the markers in the linkage maps used to aid genome assembly. Contigs were anchored to their map position by an average of ten markers, with only three contigs not supported by at least three markers (Supp. 5.7), providing support for their correct placement and therefore the non-syntenic nature of the *TPS* clusters.

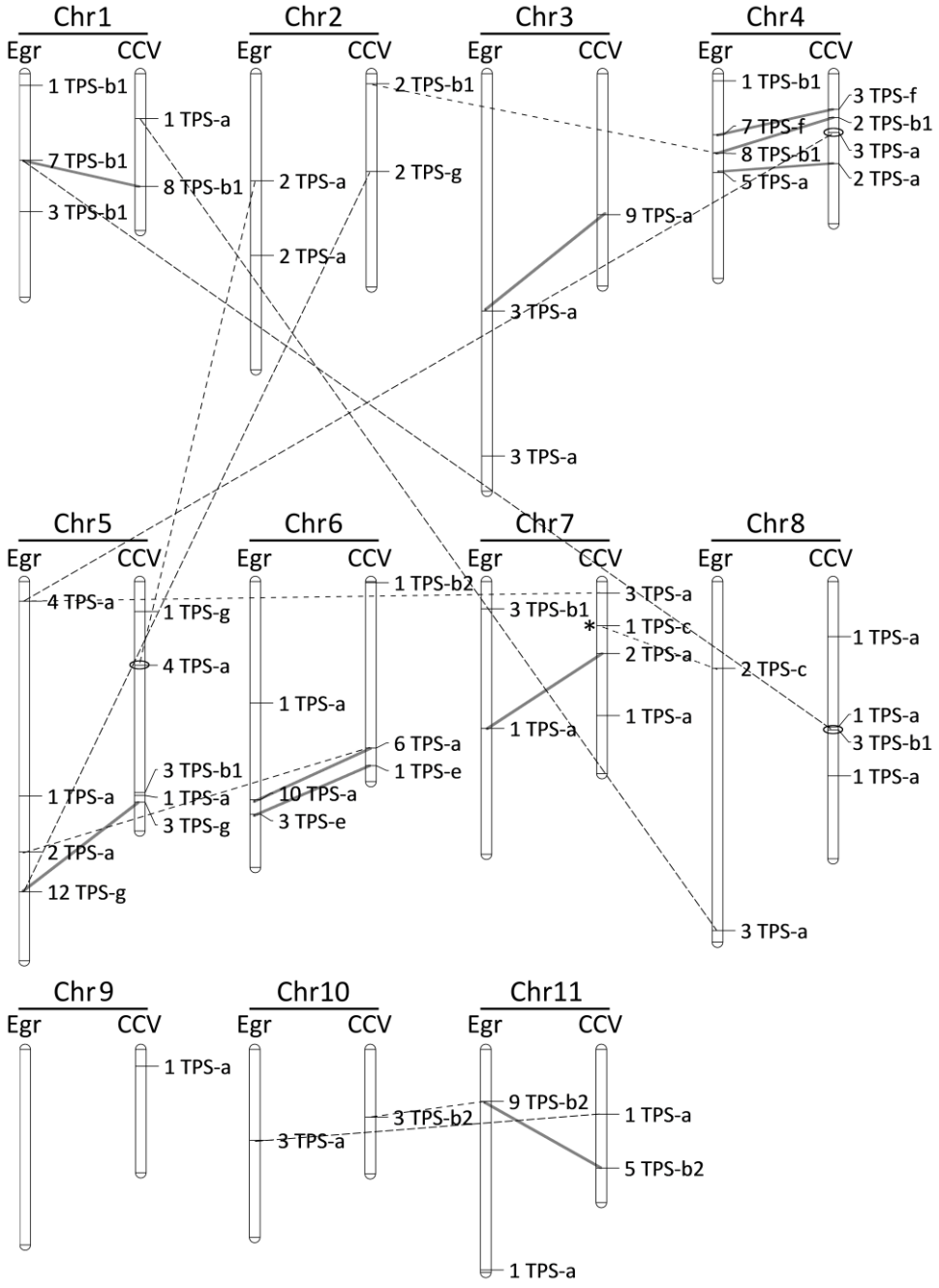


Figure 5.4. Comparison of copy number and genomic location of *TPS* physical clusters between *E. grandis* (Egr) and *C. citriodora* subsp. *variegata* (CCV). Chromosomes are scaled by physical size. Locus names show the number of *TPS* genes and the subfamily they belong to. Separate clusters on the same chromosome were defined based on both physical distance and phylogenetic relatedness (see Supp. 5.5). Solid lines indicate clusters that are both homologous and syntenic between the two species, while broken lines indicate homologous clusters that are present on different chromosomes in each species. For example, in the *TPS-b* subfamily, a cluster of eight *TPS* genes are present on chromosome 4 in *E. grandis*, in contrast to the syntenic and non-syntenic homologous clusters present in *C. citriodora* subsp. *variegata* on chromosome 4 and 2, respectively. Non-syntenic loci between *C. citriodora* subsp. *variegata* and *E. grandis* are circled (only on CCV) to indicate support for this placement based on the CCV54 genome assembly. Similarly, loci are tagged with an asterisk on CCV to indicate disagreement (see Supp. 5.6). *TPS* clusters without lines indicate that their homolog is present in the minor scaffolds of the other species and cannot be examined for synteny. Homology of singleton *TPS* genes is not shown.

Gene structure in the *TPS-a*, *-b* and *-g* subfamilies (involved in secondary metabolite synthesis) was highly conserved (Figure 5.5), with most having seven exons, and only a small proportion departing from this structure with between four and six exons. The conserved catalytic motif DDxxD (Hosfield et al. 2004; Gao et al. 2012) was generally located on the fourth exon, similar to *E. grandis* (Külheim et al. 2015). The placement of this motif on different exons was always associated with uncommon exon number. The genes from *TPS* subfamilies *-c*, *-e*, and *-f* (involved in primary metabolite synthesis) had between 10 and 13 exons, with the exception of *CorciTPS092* with six exons. The DDxxD motif in these subfamilies, when present, was not found in a consistent position. High variability was noted in the size of the first intron across all subfamilies, similar to that observed in *E. grandis* (Külheim et al. 2015). Genes ranged in size from 1,564 - 7,747 bp, with final products ranging from 337 - 739 amino acids in length (Supp. 5.1).

TPS gene expression

A heat map showing relative transcript abundance in five tissues is shown in Figure 5.6. Several expression clusters were observed, with the first expressed in both unexpanded and expanded leaves. This cluster mostly comprised genes from the *TPS-a* and *TPS-b2* subfamilies. The next cluster was characterised by expression of *TPS-a* and *TPS-b1* genes in leaves and flowers. A final cluster consisted of *TPS-a* and *TPS-b1* genes expressed in flower initials and flower buds. Of the genes involved in primary metabolism, *CorciTPS088* (*TPS-c*) was moderately expressed in bark, while CorciTPS089 (*TPS-e*) was moderately expressed across all five libraries examined. No expression was detected in the *TPS-f* subfamily.

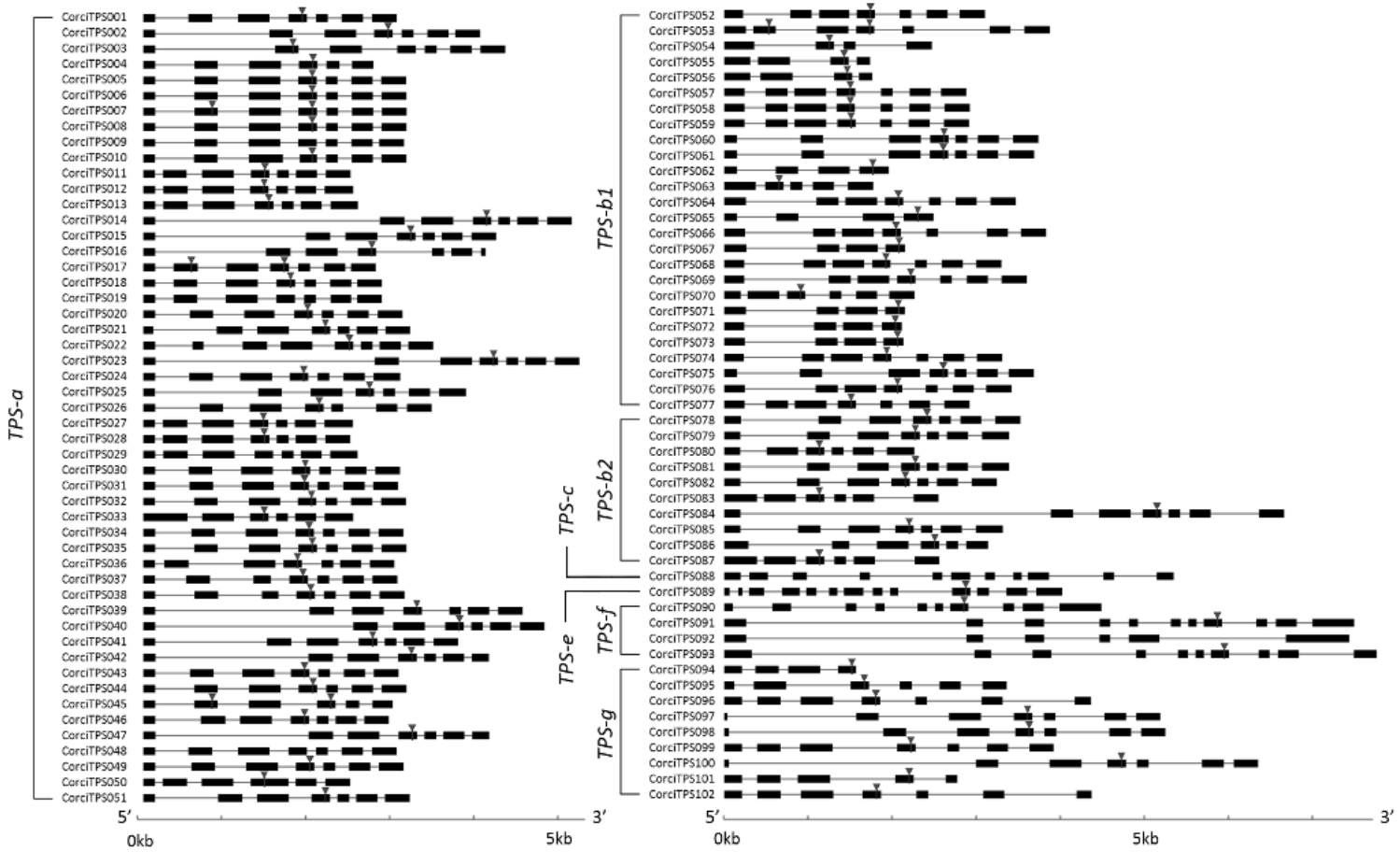


Fig. 5.5 Gene structure of the 102 putative functional *TPS* genes from *Corymbia citriodora* subsp. *variegata*. Exons are shown as boxes, while introns are shown as lines. The arrow indicates the position of the conserved DDxxD motif.

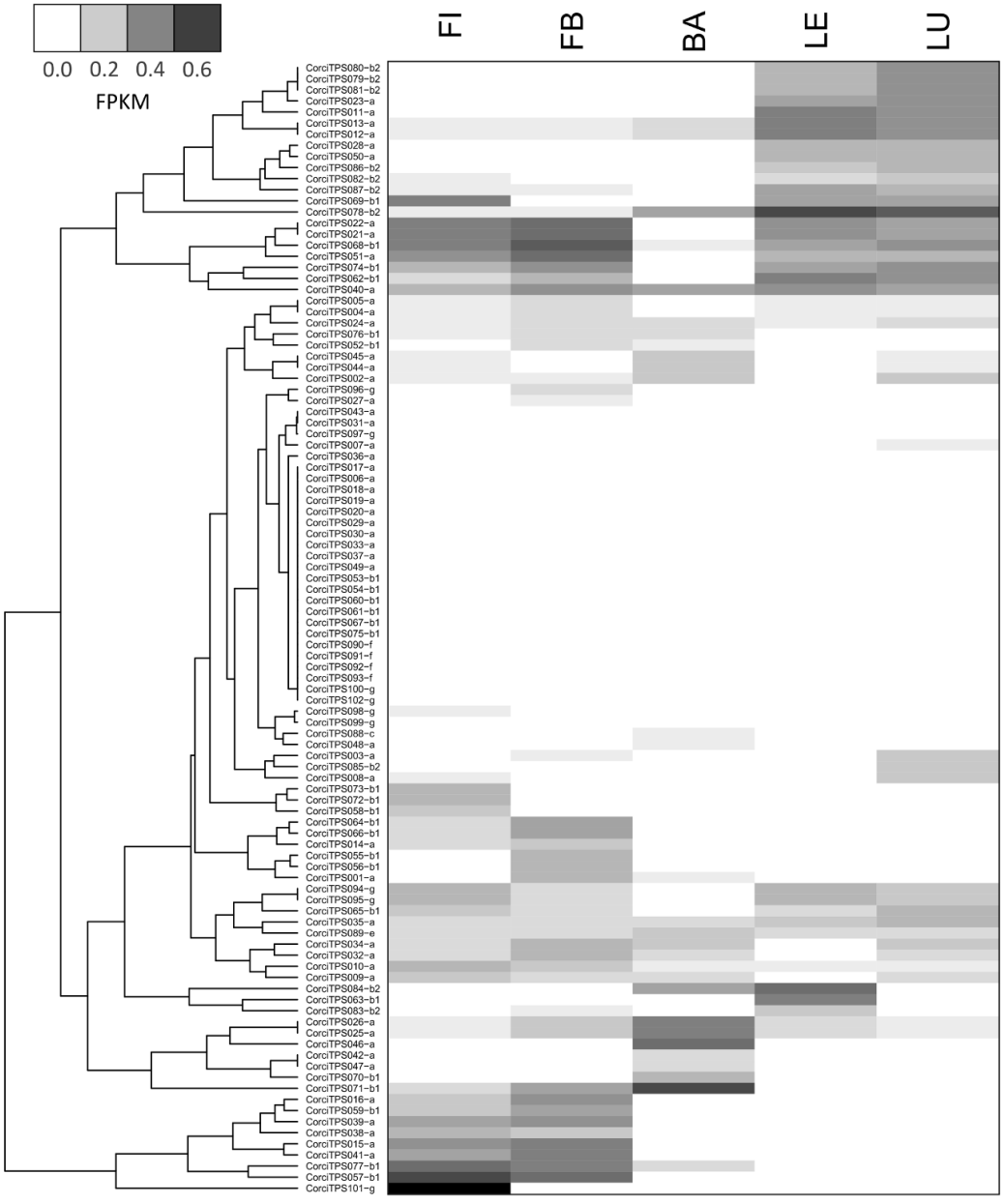


Fig. 5.6 Gene expression clustering of 102 *TPS* genes from *Corymbia citriodora* subsp. *variegata* expressed in five tissues. RNAseq data is shown as fragments per kilobase of transcript per million mapped reads (FPKM), with FPKM values normalised within libraries (largest FPKM value set to 1, with other scores scaled accordingly). The *TPS* subfamily is indicated by suffix after the gene name. The sampled tissues are: flower initials (FI), flower buds (FB), bark (BA), expanded leaf (LE), and unexpanded leaf (LU).

Discussion

Broad conservation in the eucalypt TPS family

Our analyses indicate broad conservation in gene numbers, subfamily representation, physical position and structure of clusters in the *TPS* gene family in *Corymbia citriodora* subsp. *variegata* (CCV) when compared to its divergent sister eucalypts *Eucalyptus grandis* and *E. globulus*. These eucalypts all have the same *TPS* subfamilies, which is expected given the evolution of these subfamilies is believed to pre-date the formation of the Myrtaceae (Keszei et al. 2010a). However, their similar gene numbers and subfamily representation was unexpected given (i) their divergence time from one another (Crisp et al. 2011; Thornhill et al. 2015) relative to their divergence time from the other species studied (*V. vinifera, P. trichocarpa* and *A. thaliana*) and (ii) the instability generally found in large gene families (Lynch 2007; Demuth and Hahn 2009).

We found 102 putative functional *TPS* genes in CCV, which is similar to the numbers found in *Eucalyptus grandis* (113) and *E. globulus* (106) (Külheim et al. 2015). The low variation in total number of *TPS* genes and proportional representation of each subfamily between *E. globulus, E. grandis* and CCV provides evidence for broad conservation of this gene family across these eucalypt lineages. This is in contrast to the other taxa examined in this study (*V. vinifera, P. trichocarpa* and *A. thaliana*), which varied extensively in the *TPS* family in gene number, subfamily presence and proportional representation (Aubourg et al. 2002; Martin et al. 2010; Irmisch et al. 2014). Few instances of gene orthology were detected between these three species or to the eucalypts, especially in the subfamilies involved in secondary metabolite synthesis. All these species are thought to have shared a common ancestor approximately 115 MYA (Wikström et al. 2001; Chaw et al. 2004), which, when considering the divergence of *Eucalyptus* and *Corymbia* at approximately 52 MYA (Crisp et al. 2011; Thornhill et al. 2015), makes the conservation observed between these divergent eucalypts notable. This leads us to suggest the *TPS* family size and structure observed is representative of eucalypts in general.

The number of *TPS* genes in all three eucalypts currently studied is notably high compared to other plants. Previous studies have revealed *TPS* gene family sizes ranging from one in the bryophyte *Physcomitrella patens* (Hayashi et al. 2006) to 57 in *V. vinifera* (Martin et al. 2010). Consistent with our relatively high estimates of gene birth in eucalypts compared with other taxa (Supp. 5.1), *Eucalyptus grandis* appears to have a gene duplication rate 3 - 5 times that of *Arabidopsis* and *Populus* but comparable rates of gene loss (Myburg et al. 2014), which may contribute to the higher *TPS* gene numbers in the eucalypts. Factors such

as physiology and longevity of these plants may play a role in determining the optimal *TPS* gene family size. For instance, plants that emit or store few terpenes generally have few *TPS* loci, such as *A. thaliana* and *P. trichocarpa*, while those that emit and store a more varied range of terpenes often contain more *TPS* genes (Külheim et al. 2015). Overabundance of terpenes can cause autotoxicity (Goodger et al. 2013), but plants able to store terpenes in trichomes or other glandular structures (Carr and Carr 1970) may escape this autotoxic effect. Indeed, eucalypts and *V. vinifera*, both characterised by diverse terpene profiles and the highest numbers of *TPS* loci in plants studied to date, have specialised storage structures such as oil glands. Longevity may also be a contributing factor. Due to their long generation time, more elaborate stress response mechanisms are required in perennial plants compared to herbaceous species (Soler et al. 2015). This may account for the expansion of gene families involved in stress responses in many perennials, as large numbers of genes provide an advantage in inducible responses such as pathogen resistance and other stressors and allow for rapid evolution in response to environmental change (Żmieńko et al. 2014; Sharma and Pandey 2015). For example, the *MYB* gene family, known to be involved in responses to biotic and abiotic stressors, is often expanded into large duplicate arrays in woody species but not in herbs (Soler et al. 2015), mirroring the discrepancy seen in *TPS* numbers between herbaceous species such as *A. thaliana* and the eucalypts.

Variation in the TPS genes specific to each eucalypt lineage

The conservatism at the subfamily level masks the variable expansion and contraction of gene numbers in orthologous clusters within subfamilies of *TPS* genes which, along with the much higher birth and death rate relative to the other taxa studied (Supp. 5.1), signals an evolutionarily dynamic gene family. While the importance of whole genome duplications in plant evolution is often emphasised (Soltis et al. 2014), equally as important are smaller scale duplications at the level of individual genes or gene families (Żmieńko et al. 2014). These smaller scale gene duplications (broadly defined as segmental duplications) occur when errors in DNA replication, recombination or repair generate a copy of a DNA segment containing one or more genes (Lynch and Conery 2000). Many duplicate genes are tandemly associated with their parent copy (tandem duplicates) or occur in 'localised' (within a few MB) regions of the genome, although non-localised inter/intra-chromosomal duplicates are found at lower frequency (Leister 2004; Myburg et al. 2014). Two copies of a gene are often superfluous and thus either may begin to accumulate mutations, resulting in one of several fates: neo-functionalization, where the mutated gene develops a new function; sub-functionalization, where the two copies of the gene split the function of the

original gene; or degeneration, where the gene is deactivated through mutations causing loss of function, often resulting in a pseudogene (Lynch and Conery 2000). The varied structures of *TPS* clusters in both *E. grandis* and CCV is indicative of the complex evolutionary history of this gene family (explored further in the section Physical structure of *TPS* gene clusters).

Many clades throughout the phylogenies show orthologous pairing between genes from *E. grandis* and *E. globulus.* In contrast CCV *TPS* genes are more divergent with the most closely related genes to the *Eucalyptus* species often in separate clades within subfamilies; consistent with the more recent divergence of *E. grandis* and *E. globulus* compared to the divergence of *Corymbia* and *Eucalyptus* (Crisp et al. 2011; Thornhill et al. 2015). Külheim et al. (2015) suggest that the similarities in *TPS* genes observed between *E. grandis* and *E. globulus* are a result of much of the evolution of this gene family occurring prior to their divergence. In contrast, the differences exhibited between *Corymbia* and *Eucalyptus* may result from the expansion or contraction of these gene clusters after their divergence. This is likely the case in clades with a large disparity in *TPS* gene number between CCV and the other eucalypts, for instance the *TPS-a* clade with 16 genes in CCV compared to six in *E. grandis* (Figure 5.1-a). Concerted evolution may have played a role in the differentiation of some members of the *TPS* gene family, obscuring the orthology between *TPS* genes from *E. grandis* and *E. globulus* compared with the related *Corymbia*. Concerted evolution is a process by which copies of genes separated by speciation grow to resemble neighbouring gene copies rather than their true orthologs from other species, through mechanisms such as ectopic gene conversion (Chen et al. 2007). This process may be acting throughout the *TPS* gene family and is probably the most parsimonious explanation for cases where a cluster is of similar size in all three species such as shown in Figure 5.1-d and Figure 5.2-c/5.2-d, as opposed to multiple instances of lineage specific expansion. For example, evidence for gene conversion was found between CCV *TPS-b2* genes in a clade with similar numbers of genes in each species (Figure 5.2-d, Supp. 5.8), lending support to this hypothesis. However, sequencing and annotation of the *TPS* gene family in a sister taxa of the eucalypts [e.g. *Arillastrum, Allosyncarpia, Stockwellia,* or *Eucalyptopsis* (Macphail and Thornhill 2016)] is needed to provide a suitable outgroup to elucidate which mode of evolution affected specific clades as well as whether expansion/contraction of clusters occurred in the *Corymbia* or *Eucalyptus* lineage.

Variation in the TPS subfamilies involved in secondary metabolite synthesis

While the overall proportional representation of each *TPS* subfamily is not significantly different, CCV and *Eucalyptus* exhibit marked differences in gene number within several physical clusters in subfamilies *TPS-a*, *-b* and *-g* (Figure 5.1, Figure 5.2). These *TPS* subfamilies are involved in the synthesis of secondary metabolites, which play roles in biotic/abiotic stress responses (Chen et al. 2011). Differential expansion/contraction of gene clusters between species has been often observed, including in the receptor kinase gene family across Brassicaceae (Hofberger et al. 2015) and the *MYB* family across various taxa (Wilkins et al. 2009; Soler et al. 2015); specifically, *R2R3-MYB* gene number varies from 118 in *V. vinifera* to 192 in *P. trichocarpa* (Wilkins et al. 2009). There is potential that localised duplication of these genes facilitates the gain of new function while keeping new copies under similar regulatory control, either through directly copying the original regulatory elements or through other controls such as shared promoters (Williams and Bowles 2004). This mechanism is thought to provide a selective advantage in inducible responses such as biotic resistance, as their shared regulatory control will express both the original and this new potentially advantageous gene when a response is induced (Leister 2004; Hanada et al. 2008). If advantageous, these duplicate genes will be maintained, leading to the expansion of clusters as seen in the *TPS* genes presented here.

Conservation in the TPS subfamilies involved in primary metabolite synthesis

In contrast to the other subfamilies, those involved in the synthesis of primary metabolites (*TPS-c*, *-e,* and to a lesser extent *-f*) are more conserved in cluster copy number across eucalypt species (Figure 5.3), likely reflecting stronger selective constraints on primary *versus* secondary metabolites (Chen et al. 2011). Conservation within a selectively constrained section of an expanded gene family has been previously observed in families such as *MYB* (Wilkins et al. 2009) and *SBP-box* in plants (Zhang et al. 2015), consistent with our findings. As well as greater conservation of gene numbers within clusters, there was also greater conservation of synteny in the subfamilies involved in the synthesis of primary metabolites than those involved in secondary metabolite synthesis across the eucalypts. All *TPS* loci involved in primary metabolite synthesis (*TPS-c*, *-e* and *-f*) were syntenic between *E. grandis* and CCV with no evidence of transposition between chromosomes (aside from a single *TPS-c* gene for which there is evidence of misassembly). The hypothesis of 'gene balance' suggests that duplicate genes that act in dosage-dependant manners are usually only retained after polyploidy events (Veitia 2004). In the event of a small scale duplication the other parts of the metabolic pathway are often unchanged, which may cause unused product to accumulate and result in detrimental dosage effects (Freeling 2009; Tang and

Amon 2013). The conservation seen in *TPS* gene families involved in primary metabolism is consistent with this hypothesis. It is interesting to note that *A. thaliana* has only a single copy of *TPS-c*, *-e* and *-f* genes, while the eucalypts generally have two or more (Table 5.1, Figure 5.3). Whole genome duplications specific to each lineage have been detected in both *Arabidopsis* and the plant order Myrtales to which the family Myrtaceae belongs (Arabidopsis Genome Initiative 2000; Myburg et al. 2014), suggesting the persistence of *TPS* duplicates in these subfamilies was not advantageous for *Arabidopsis*.

Contributions of stochastic and selective pressures to the variation in the TPS gene family

The balance observed in total subfamily representation may be due to the stochastic nature of mechanisms driving gene duplication and loss (Lynch 2007). While selection will act to fix or purify (inactivate) a beneficial or detrimental duplicate gene, these duplicates can also be selectively neutral, leading to their maintenance and subsequent cluster expansion with very minor impact on the fitness of the organism (Iskow et al. 2012). Maintaining a large library of neutral genes can be selectively advantageous in areas of environmental volatility, allowing the organism to be potentially 'pre-adapted' to stressors (Hurles 2004; Hanada et al. 2008; Kondrashov 2012). Genes in large families have also been shown to be gained and lost at very similar rates through analysis of gene 'birth and death' across gene families in multiple genomes (Demuth and Hahn 2009; Szöllősi and Daubin 2012) and specifically in *A. thaliana* (Cannon et al. 2004). If changes in cluster number are occurring across the entire *TPS* gene family within a species, expansion in one cluster may be countered by degeneration in another, contributing to the overall balance in subfamily representation despite the apparent species specific gain and loss of loci observed between the eucalypts.

The conservation of high *TPS* numbers and subfamily proportional representation across the eucalypts despite the extensive variation in some subfamilies may be a signal that selection is involved. While selection may be acting on the phenotype to drive duplicated genes to fixation or degeneration, the combined effect of these large gene families is also likely to be influenced by selection. The maintenance of a large library of genes, while advantageous in some situations (Żmieńko et al. 2014; Sharma and Pandey 2015), may have associated costs. These include increasing expression and regulation requirements with increasing number (Schiffer et al. 2016) and the possibility of 'runaway expansion' contributing to genome instability (Gijzen 2009; Schiffer et al. 2016), which may be detrimental enough to select against further expansion of *TPS* clusters. Given that increased gene copies often result in increased expression of the subsequent product, there may also be a maximum amount of *TPS* genes that eucalypts can support without experiencing specific deleterious gene dosage effects. For example, the overexpression of particular *TPS-*

a genes has been shown to retard growth in tomato (Fray et al. 1995), tobacco (Busch et al. 2002) and *A. thaliana* (Aharoni et al. 2003; Ee et al. 2014) (though not without exception, see Schnee et al. (2006)). This is thought to be the result of under-expression of primary metabolites, such as gibberellin and abscisic acid, due to terpenoid precursor reserves becoming exhausted by the over-synthesis of secondary metabolites. This theory, along with the autotoxicity explored earlier, may select against unregulated expansion in *TPS* clusters and contribute to the stability in *TPS* gene numbers and subfamily representation across the eucalypts.

Pseudogenes and expression of TPS loci

The most likely fate of duplicate genes is to be released from selection and acquire mutations which render them non-functional, resulting in pseudogenes. We found 25 *TPS* pseudogenes in the CCV genome, which is 24.5% of the total putative *TPS* gene family size (Supp. 5.2). Fifteen of these occurred in the main chromosome assemblies of which all but one was within existing TPS clusters, providing further evidence for the extensive history of local gene duplications in this lineage. Of the 25 pseudogenes, 15 showed evidence of expression, which is an interesting finding. While pseudogenes are thought to play at most a passive role in the genome, such as being sequence donor/receptors for proximal genes (Zheng and Gerstein 2007), it has been shown that some pseudogene transcripts in humans can bind to mRNA from related functional genes and affect their expression (Vinckenbosch et al. 2006). If this is the case in *Corymbia,* these pseudogenes may be part of a mechanism for modulating gene expression.

Expression analysis of putative functional genes (Figure 5.6) revealed several distinct expression clusters each of which involved multiple subfamilies. Secondary metabolite subfamilies were represented across most tissues, consistent with the broad applications of these terpenoids (Keszei et al. 2010a). An interesting pattern was detected in the *TPS-b2*, which were highly expressed in the unexpanded and expanded leaf tissue libraries, with low expression in other libraries. This subfamily is involved in the synthesis of isoprene, a terpenoid hypothesized to confer thermotolerance (Peñuelas et al. 2005). Isoprene is known to lower tissue surface temperature when emitted (Sasaki et al. 2007) and also improves the stability of plant membranes (Singsaas et al. 1997). As these both affect photosynthetic rate, the higher expression of isoprene in leaf tissue is consistent with these modes of action. The analysis also showed the *TPS-f* subfamily in CCV was not expressed in the five tissues examined (flower buds and initials, unexpanded and expanded leaf, and bark). This is consistent with similar analysis in *E. grandis,* which revealed that most *TPS-f* genes were solely expressed in root tissue (Külheim et al. 2015), a tissue not covered by our

analysis. This subfamily also showed higher divergence than the other primary metabolite subfamilies in all three eucalypt species. Due to this non-typical expression pattern, Külheim et al. (2015) suggest *TPS-f* play a role mediating interactions with herbivores and other soil organisms (Wenke et al. 2010), or influencing allelopathic effects (del Moral and Muller 1970). Indeed, the divergence of the *TPS-f* subfamily across the eucalypts could signal the potential environmental specificity of these interactions.

Physical structure of TPS gene clusters

In both *E. grandis* and CCV, most *TPS* genes were clustered in localised regions of the genome (spanning up to 3.5 MB). The fact that each cluster contained only *TPS* genes from the same subfamily that are also closely related in sequence (with the exception of one *TPS-a* gene located within a dispersed *TPS-b* cluster on chromosome five of CCV) suggests they were generated by localised or tandem gene duplication. Indeed, *E. grandis* is characterised by a high rate of tandem duplication relative to other plants (Myburg et al. 2014), which has been proposed as the main reason for the extensive *TPS* family in the eucalypts (Myburg et al. 2014; Külheim et al. 2015), as well as gene families in many other species (Kliebenstein et al. 2001; Leister 2004; Hofberger et al. 2013; Hofberger et al. 2015; Li et al. 2015). In some cases the eucalypt *TPS* genes were in true tandem arrays with no genes interspersed (the largest being a syntenic cluster with six *TPS-a* genes in CCV and 10 in *E. grandis*), while in most cases several non-*TPS* genes were present within these clusters, ranging from 1 - 192 genes separating the closest *TPS* in CCV (Table 5.2). The varying spans and intervening gene number of *TPS* clusters in *E. grandis* and CCV likely reflect the many ways clusters can form and be subsequently rearranged (Leister 2004; Lynch 2007; Field et al. 2011). For example, segmental duplications range in size and can result in partial genes to large-scale genome segments being copied and translocated to new inter-/intra-chromosomal positions (Flagel and Wendel 2009; Wang et al. 2012), or positions local to the origin (Cannon et al. 2004). Hence, the duplication process may initially result in tandem, localised or dispersed gene pairs. Superimposed on this variation, localised (and tandem) duplications can be subsequently dispersed by various mechanisms of genome rearrangement (Lynch 2007; Field et al. 2011); including inversions, insertion/deletions, translocations and further segmental duplications (the tandem expansion of *NB-LRR* genes contained within several *TPS* clusters may be an example of the latter, see Supp. 5.5). Differentiating between these various processes requires determining the relative age of duplications, which due to the inherent difficulties introduced by concerted evolution obfuscating mutations is beyond the scope of this study (Mendivil-Ramos and Ferrier 2012).

The physical clustering of non-homologous, but functionally related genes is an emerging theme in plant genomics, particularly in the case of secondary metabolite pathways (Chu et al. 2011; Field et al. 2011; Takos and Rook 2012). Many non-*TPS* genes within *TPS* clusters have putative functions that may interact with or complement the function of *TPS* genes (Supp. 5.5). Genes potentially involved in the synthesis of terpene precursors, such as prenyl transferases, along with those involved in post-translational modification of terpenes such as cytochrome c oxidases and NAD-dependant dehydrogenases (Keszei et al. 2008) were found within *TPS* clusters in both *E. grandis* (Külheim et al. 2015) and CCV. Also found were genes from the *NB-LRR*, *MYB* and *WRKY* families, which among other things are involved in pest resistance (Liu et al. 2004; Eitas and Dangl 2010), much like *TPS* genes. The location of these genes within *TPS* clusters may be advantageous, as genes involved in the same biosynthetic pathway or in similar responses can be regulated together at the chromatin level (Field and Osbourn 2008; Chu et al. 2011). This arrangement may also be beneficial for inheritance, as a collection of beneficial alleles from a single metabolic pathway are less likely to be separated by recombination when in close proximity (Chu et al. 2011).

Conclusions

This study contributes to a greater understanding of the terpene synthase gene family through detailed annotation in the recently assembled *C. citriodora* subsp. *variegata* genome and comparative analysis with the previously studied *E. grandis* and *E. globulus*. These *Eucalyptus* species have the most *TPS* loci discovered in any plant to date, and our results show the large size of this gene family is conserved in the sister genus *Corymbia*, suggesting this may be a characteristic of the eucalypts. Both the proportional representation of subfamilies and the syntenic physical position of gene clusters indicated a high degree of conservation in the *TPS* gene family between CCV and *E. grandis.* Despite this conservation, cluster specific variation within subfamilies involved in secondary metabolite synthesis were observed, and we discuss the potential contributions of selection, concerted evolution and stochastic processes to this observation. The higher degree of conservation of *TPS* genes involved in primary metabolite synthesis is likely due to greater selective constraints.

Supplementary material

Supp. 5.1 *C. citriodora* subsp. *variegata* putative functional terpene synthase genes from the CCV18 genome assembly.

The sequence column has been filtered out, as it is unfeasible to present in this thesis due to its length. The full version of this table can be found alongside the published paper (referred to as Table S1) at:
https://www.nature.com/articles/s41437-018-0058-1

Overlapping model	Gene name	Chromosome/ scaffold	Subfamily	Start 5'	End 3'		Length (cDNA)	Sequence (see note)	FS/PS[a]	FB[b]	FI	UL	EL	B	Class
CorciG021131	CorciTPS001	chr1	a	9,037,469	9,040,474	+	1,701		0	y	n	n	n	y	1
CorciG014645	CorciTPS002	chr3	a	27,596,209	27,600,202	+	1,698		0	y	y	y	n	y	1
CorciG014654	CorciTPS003	chr3	a	27,818,837	27,823,128	+	1,698		1	y	n	y	n	n	2
CorciG014655	CorciTPS004	chr3	a	27,834,442	27,837,169	+	1,419		0	y	y	y	y	y	1
CorciG014655	CorciTPS005	chr3	a	27,840,600	27,843,715	+	1,698		0	y	y	y	y	y	1
NA	CorciTPS006	chr3	a	28,227,718	28,230,838	+	1,698		0	n	n	n	n	n	3
NA	CorciTPS007	chr3	a	28,239,510	28,242,634	+	1,698		0	n	y	y	y	n	1
NA	CorciTPS008	chr3	a	28,253,723	28,256,843	+	1,698		0	y	y	y	y	y	1
CorciG014678	CorciTPS009	chr3	a	28,268,386	28,271,477	+	1,692		0	y	y	y	n	y	1
CorciG014846	CorciTPS010	chr3	a	31,195,150	31,192,031	-	1,704		1	y	y	y	y	y	2
NA	CorciTPS011	chr4	a	11,819,029	11,816,572	-	1,692		1	n	n	y	y	n	2
CorciG023522	CorciTPS012	chr4	a	11,825,354	11,822,866	-	1,752		0	y	y	y	y	y	1
CorciG023522	CorciTPS013	chr4	a	11,831,277	11,828,732	-	1,740		0	y	y	y	y	y	1
CorciG023936	CorciTPS014	chr4	a	18,093,425	18,088,344	-	1,704		0	y	y	n	n	n	1
CorciG023938	CorciTPS015	chr4	a	18,112,220	18,108,037	-	1,704		0	y	y	n	n	n	1
CorciG010840	CorciTPS016	chr5	a	15,915,747	15,919,805	+	1,467		0	y	y	n	n	n	1
CorciG010908	CorciTPS017	chr5	a	17,107,776	17,105,019	-	1,701		1	n	n	n	n	n	3

CorciG010908	CorciTPS018	chr5	a	17,112,248	17,109,418	-	1,701	0	n	n	n	n	n	1
CorciG010909	CorciTPS019	chr5	a	17,135,597	17,132,769	-	1,692	0	n	n	n	n	n	3
CorciG012362	CorciTPS020	chr5	a	43,015,800	43,012,727	-	1,674	1	n	n	n	n	n	3
CorciG006219	CorciTPS021	chr6	a	33,411,585	33,408,422	-	1,695	1	y	y	y	y	y	2
CorciG006219	CorciTPS022	chr6	a	33,432,240	33,428,805	-	1,842	0	y	y	y	y	y	1
CorciG006220	CorciTPS023	chr6	a	33,467,016	33,461,846	-	1,704	0	n	n	y	y	n	1
CorciG006221	CorciTPS024	chr6	a	33,482,436	33,485,482	+	1,719	1	y	y	y	y	y	2
CorciG006222	CorciTPS025	chr6	a	33,500,561	33,496,735	-	1,716	1	y	y	y	y	y	2
CorciG006222	CorciTPS026	chr6	a	33,515,351	33,511,933	-	1,692	0	y	y	y	y	y	1
CorciG015778	CorciTPS027	chr7	a	1,778,430	1,780,915	+	1,749	0	y	y	n	n	n	1
CorciG015836	CorciTPS028	chr7	a	2,568,513	2,566,060	-	1,707	0	n	y	y	y	n	1
CorciG015838	CorciTPS029	chr7	a	2,587,681	2,585,141	-	1,737	0	n	n	n	n	n	3
CorciG016570	CorciTPS030	chr7	a	13,582,561	13,585,605	+	1,695	2	n	n	n	n	n	3
CorciG016674	CorciTPS031	chr7	a	15,343,940	15,340,920	-	1,701	0	n	n	y	n	n	1
CorciG017329	CorciTPS032	chr7	a	26,940,768	26,943,881	+	1,692	0	y	y	y	y	y	1
CorciG000907	CorciTPS033	chr8	a	11,042,040	11,044,528	+	1,845	0	n	n	n	n	n	3
CorciG001987	CorciTPS034	chr8	a	29,233,006	29,236,090	+	1,698	0	y	y	y	y	y	1
CorciG002634	CorciTPS035	chr8	a	39,075,938	39,079,058	+	1,698	0	y	y	y	y	y	1
CorciG026984	CorciTPS036	chr9	a	3,364,546	3,367,524	+	1,701	0	n	y	n	n	n	1
CorciG019224	CorciTPS037	chr11	a	13,004,369	13,007,381	+	1,530	0	n	n	n	n	n	3
CorciG029535	CorciTPS038	scaffold291	a	362,842	359,746	-	1,521	0	y	y	n	n	n	1
CorciG029892	CorciTPS039	scaffold366	a	224,590	229,087	+	1,704	0	y	y	n	n	n	1
CorciG029897	CorciTPS040	scaffold366	a	309,579	304,822	-	1,704	0	y	y	y	y	y	1
CorciG031211	CorciTPS041	scaffold930	a	123,042	126,774	+	1,704	0	y	y	n	n	n	1
CorciG030952	CorciTPS042	scaffold1110	a	14,816	18,912	+	1,614	1	n	n	n	n	y	2
CorciG031972	CorciTPS043	scaffold1321	a	92,883	89,857	-	1,701	0	n	n	y	n	n	1
CorciG031126	CorciTPS044	scaffold1391	a	8,166	5,045	-	1,698	0	y	y	y	n	y	1

CorciG031126	CorciTPS045	scaffold1391	a	15,184	12,227	-	1,410	1	y	y	y	n	y	2
CorciG031505	CorciTPS046	scaffold1449	a	18,216	15,304	-	1,704	0	n	n	n	n	y	1
CorciG031507	CorciTPS047	scaffold1449	a	31,766	27,666	-	1,617	0	n	n	n	n	y	1
CorciG031449	CorciTPS048	scaffold1523	a	24,880	21,876	-	1,698	1	y	n	n	n	y	2
CorciG033462	CorciTPS049	scaffold1835	a	45,877	48,964	+	1,698	0	n	n	n	n	n	3
CorciG031890	CorciTPS050	scaffold2657	a	16,857	14,405	-	1,794	0	n	n	y	y	n	1
CorciG034389	CorciTPS051	scaffold3132	a	10,199	7,041	-	1,698	1	y	y	y	y	n	2
CorciG021943	CorciTPS052	chr1	b1	21,355,181	21,352,086	-	1,737	0	y	n	n	n	y	1
NA	CorciTPS053	chr1	b1	22,582,107	22,585,972	+	1,800	0	n	n	n	n	n	3
NA	CorciTPS054	chr1	b1	22,598,557	22,601,020	+	1,011	0	n	n	n	n	n	3
NA	CorciTPS055	chr1	b1	22,620,399	22,622,130	+	1,062	0	y	y	n	n	n	1
NA	CorciTPS056	chr1	b1	22,631,710	22,633,467	+	1,062	0	y	n	n	n	n	1
NA	CorciTPS057	chr1	b1	23,186,987	23,189,860	+	1,791	0	y	y	n	n	n	1
NA	CorciTPS058	chr1	b1	23,199,888	23,202,803	+	1,785	0	n	y	n	n	n	1
NA	CorciTPS059	chr1	b1	23,215,946	23,218,853	+	1,785	0	y	y	n	n	n	1
CorciG006911	CorciTPS060	chr2	b1	2,024,189	2,020,461	-	1,701	0	n	n	n	n	n	3
CorciG006911	CorciTPS061	chr2	b1	2,034,153	2,030,474	-	1,701	0	n	n	n	n	n	3
CorciG023301	CorciTPS062	chr4	b1	8,776,542	8,778,497	+	1,143	0	y	y	y	y	n	1
CorciG023307	CorciTPS063	chr4	b1	8,944,409	8,946,178	+	1,284	0	n	n	n	y	n	1
CorciG012193	CorciTPS064	chr5	b1	40,058,179	40,061,640	+	1,800	0	y	y	y	n	n	1
CorciG012384	CorciTPS065	chr5	b1	43,383,395	43,385,879	+	1,143	0	y	y	y	y	n	1
CorciG012407	CorciTPS066	chr5	b1	43,846,108	43,842,285	-	1,791	0	y	y	n	n	n	1
NA	CorciTPS067	chr8	b1	29,775,806	29,777,953	+	1,134	0	n	n	n	n	n	3
NA	CorciTPS068	chr8	b1	29,787,754	29,791,045	+	1,782	0	y	y	y	y	y	1
CorciG002026	CorciTPS069	chr8	b1	29,819,611	29,823,204	+	1,782	0	n	y	y	y	n	1
CorciG029544	CorciTPS070	scaffold429	b1	40,895	43,153	+	1,473	1	n	n	n	n	y	2
CorciG030143	CorciTPS071	scaffold494	b1	96,344	98,487	+	1,128	0	y	y	y	n	y	1

CorciG030841	CorciTPS072	scaffold1014	b1	70,235	72,343	+	1,092	0	n	y	n	n	n	1
CorciG030847	CorciTPS073	scaffold1014	b1	143,780	145,908	+	1,119	0	y	y	n	n	n	1
CorciG032045	CorciTPS074	scaffold1042	b1	42,932	39,633	-	1,773	0	y	y	y	y	n	1
CorciG032944	CorciTPS075	scaffold1784	b1	30,877	27,200	-	1,701	0	n	n	n	n	n	3
CorciG033988	CorciTPS076	scaffold2350	b1	15,337	18,746	+	1,788	0	y	y	y	y	y	1
CorciG034186	CorciTPS077	scaffold3281	b1	8,377	5,466	-	1,785	0	y	y	n	n	y	1
CorciG003747	CorciTPS078	chr6	b2	197,518	194,003	-	1,746	0	y	y	y	y	y	1
NA	CorciTPS079	chr10	b2	13,586,330	13,589,711	+	1,755	2	n	n	y	y	n	2
NA	CorciTPS080	chr10	b2	13,599,324	13,601,580	+	1,497	0	n	n	y	y	n	1
NA	CorciTPS081	chr10	b2	13,621,094	13,624,474	+	1,761	0	n	n	y	y	n	1
CorciG019951	CorciTPS082	chr11	b2	23,599,751	23,602,986	+	1,737	0	y	y	y	y	n	1
CorciG019952	CorciTPS083	chr11	b2	23,614,520	23,617,062	+	1,671	1	y	y	n	y	n	2
CorciG019977	CorciTPS084	chr11	b2	23,938,898	23,945,543	+	1,734	0	n	n	n	y	y	1
NA	CorciTPS085	chr11	b2	23,997,474	24,000,781	+	1,758	1	n	y	n	n	n	2
CorciG019978	CorciTPS086	chr11	b2	24,017,408	24,020,539	+	1,482	0	y	y	y	y	n	1
CorciG035333	CorciTPS087	scaffold3301	b2	9,457	6,903	-	1,671	1	y	y	y	y	n	2
CorciG016305	CorciTPS088	chr7	c	8,874,051	8,879,385	+	1,836	2	n	n	n	n	y	2
CorciG006508	CorciTPS089	chr6	e	37,020,792	37,024,804	+	2,082	0	y	y	y	y	y	1
CorciG023161	CorciTPS090	chr4	f	6,947,577	6,943,102	-	1,863	0	n	n	n	n	n	3
CorciG023192	CorciTPS091	chr4	f	7,312,240	7,304,764	-	2,217	0	n	n	n	n	n	3
CorciG023192	CorciTPS092	chr4	f	7,328,676	7,321,257	-	1,923	0	n	n	n	n	n	3
CorciG033231	CorciTPS093	scaffold2655	f	10,033	17,780	+	1,944	0	n	n	n	n	n	3
NA	CorciTPS094	chr2	g	19,644,329	19,642,765	-	1,080	1	y	y	y	y	n	2
CorciG008084	CorciTPS095	chr2	g	19,672,177	19,668,823	-	1,401	0	y	y	y	y	n	1
CorciG010170	CorciTPS096	chr5	g	6,021,109	6,016,757	-	1,671	0	y	y	y	n	n	1
CorciG012434	CorciTPS097	chr5	g	44,291,023	44,296,200	+	1,587	0	n	n	y	n	n	1
CorciG012436	CorciTPS098	chr5	g	44,361,036	44,366,272	+	1,605	0	n	y	y	n	y	1

CorciG012437	CorciTPS099	chr5	g	44,387,484	44,391,394	+	1,767	0	n	y	y	n	y	1
CorciG031464	CorciTPS100	scaffold1656	g	28,436	34,772	+	1,605	0	n	n	n	n	n	3
CorciG031466	CorciTPS101	scaffold1656	g	57,028	59,795	+	1,230	0	n	y	n	n	n	1
CorciG033373	CorciTPS102	scaffold2152	g	7,887	12,247	+	1,656	0	n	n	n	n	n	3

[a] Frame shifts or premature stop codons

[b] Presence or absence of expression in various libraries. FB = Flower buds, FI = flower initials, UL = unexpanded leaf, EL = expanded leaf, B = bark.

Supp. 5.2 *C. citriodora* subsp. *variegata* terpene synthase pseudogenes from the CCV18 genome assembly

Overlapping model	Chromosome/scaffold	Subfamily	Start 5'	End 3'	Orientation	Length	FB	FI	UL	EL	B	Class
CorciG022178	chr1	a	25,193,230	25,190,744	-	2,486	y	y	y	y	y	4
CorciG023521	chr4	a	11,795,521	11,792,984	-	2,537	n	y	n	y	n	4
CorciG016573	chr7	a	13,640,872	13,643,916	+	3,044	n	n	n	n	n	4
CorciG002632	chr8	a	38,991,197	38,994,317	+	3,120	y	y	y	y	y	4
CorciG029891	scaffold366_size394434	a	209,151	212,818	+	3,667	y	n	n	n	y	4
CorciG029894	scaffold366_size394434	a	256,823	259,100	+	2,277	y	n	n	n	y	4
CorciG034645	scaffold4257_size7774	a	4,253	7,294	+	3,041	n	n	n	n	n	4
CorciG034669	scaffold2130_size38350	a	12,929	9,262	-	3,667	n	n	n	n	n	4
CorciG014676	chr3	a	28,221,381	28,224,290	+	2,909	y	y	y	y	y	4
CorciG000907	chr8	a	11,030,435	11,032,975	+	2,540	n	n	n	n	n	4
NA	chr3	a	28,247,575	28,250,305	+	2,730	y	y	y	n	y	4
CorciG014675	chr3	a	28,184,428	28,187,531	+	3,103	n	n	n	n	n	4
CorciG031888	scaffold2657_size49423	a	6,603	4,288	-	2,315	n	n	n	n	n	4
CorciG035241	scaffold2783_size22306	a	6,464	4,084	-	2,380	y	n	y	y	y	4
NA	chr8	a	3,543,086	3,544,120	+	1,034	n	n	n	n	n	4
CorciG023299	chr4	b	8,737,927	8,741,333	+	3,406	y	y	y	y	y	4
CorciG034480	scaffold3769_size11075	b	3,659	2	-	3,657	n	n	n	n	n	4
NA	chr1	b	22,575,468	22,577,940	+	2,472	y	n	y	n	y	4
NA	scaffold1218_size103965	b	70,850	71,727	+	877	y	y	n	n	n	4
CorciG019956	chr11	b2	23,658,990	23,661,484	+	2,494	y	y	y	y	y	4
NA	chr11	b2	23,992,914	23,995,934	+	3,020	n	n	n	n	n	4
CorciG023191	chr4	f	7,268,615	7,263,468	-	5,147	n	n	n	n	y	4
CorciG023162	chr4	f	6,965,238	6,960,567	-	4,671	n	y	y	y	y	4
CorciG031466	scaffold1656_size82306	g	71,677	74,958	+	3,281	n	y	y	n	y	4
CorciG031461	scaffold1656_size82306	g	7,814	10,093	+	2,279	n	n	n	n	n	4

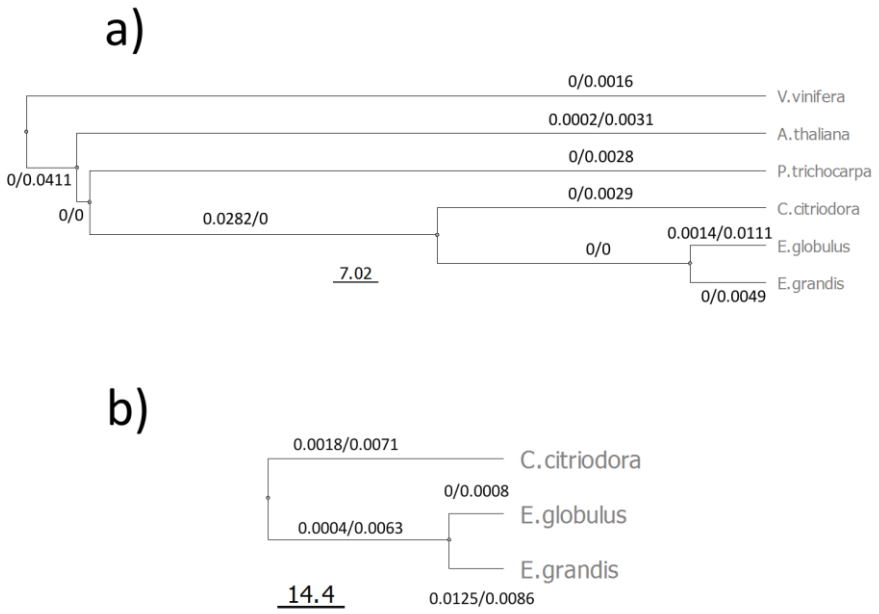
Supp. 5.3 *C. citriodora* subsp. *variegata* terpene synthase genes from the CCV54 genome assembly

The sequence column has been filtered out, as it is unfeasible to present in this thesis due to its length. The full version of this table can be found alongside the published paper (referred to as Table S1) at:
https://www.nature.com/articles/s41437-018-0058-1

Chromosome/ scaffold	Subfamily	Start 5'	End 3'		Length (DNA)	Sequence (see note)	Designation
2	a	17,017,882	17,021,989	+	1,617		Putative
2	a	17,044,535	17,047,438	+	1,704		Putative
3	a	20,994,491	20,993,220	-	900		Putative
4	a	10,981,390	10,986,481	+	1,674		Putative
5	a	6,303,457	6,306,577	+	1,698		Putative
6	a	20,208,453	20,204,864	-	1,698		Putative
6	a	20,220,698	20,218,318	-	1,497		Putative
6	a	20,259,130	20,255,304	-	1,716		Putative
6	a	20,271,202	20,267,783	-	1,692		Putative
6	a	21,100,237	21,103,228	+	1,692		Putative
6	a	21,112,913	21,109,515	-	1,692		Putative
6	a	22,908,109	22,902,346	-	1,338		Putative
8	a	10,890,726	10,887,475	-	1,362		Pseudogene
8	a	10,909,970	10,905,513	-	1,692		Putative
8	a	10,921,556	10,919,041	-	1,542		Putative
8	a	11,134,231	11,130,209	-	1,692		Putative
10	a	6,555,242	6,558,226	+	1,701		Putative
11	a	6,712,959	6,715,895	+	1,698		Putative
11	a	8,329,572	8,326,560	-	1,698		Putative
scaffold2013	a	8,623	11,654	+	1,701		Putative
scaffold2193	a	27,363	30,999	+	1,242		Putative
scaffold2489	a	160,715	158,250	-	1,545		Putative
scaffold2720	a	42,607	41,943	-	558		Pseudogene
scaffold2865	a	22,875	16,708	-	1,662		Putative
scaffold2865	a	24,283	21,743	-	1,734		Putative
scaffold4275	a	21,356	18,872	-	1,767		Pseudogene
scaffold4762	a	7,596	5,103	-	1,539		Putative
scaffold4836	a	310	2,195	+	1,242		Putative
scaffold653	a	61,544	64,663	+	1,281		Putative
scaffold6876	a	16,644	13,886	-	1,698		Putative
scaffold8693	a	2,738	467	-	1,704		Putative
scaffold90	a	181,034	178,902	-	1,575		Putative
1	b	8,440,439	8,436,927	-	1,125		Putative
1	b	13,530,921	13,529,193	-	1,773		Putative
1	b	13,539,518	13,537,783	-	1,059		Putative
1	b	13,550,212	13,546,450	-	1,062		Pseudogene
1	b	13,556,925	13,554,864	-	1,635		Putative
1	b	14,225,359	14,229,245	+	978		Putative

1	b	15,274,552	15,271,581	-	1,815	Putative
5	b	28,376,067	28,379,891	+	1,806	Putative
8	b	24,735,659	24,733,843	-	1,791	Putative
8	b	24,748,219	24,744,835	-	888	Putative
scaffold10090	b	746	1,779	+	1,773	Putative
scaffold430	b	78,348	81,393	+	861	Putative
scaffold5412	b	17,041	19,169	+	1,485	Pseudogene
scaffold6330	b	9,965	13,559	+	1,341	Putative
scaffold8556	b	746	2,477	+	1,782	Putative
4	b2	5,544,305	5,541,466	-	1,062	Putative
4	b2	5,569,318	5,565,639	-	1,560	Putative
11	b2	16,223,864	16,229,351	+	1,701	Putative
scaffold1409	b2	69,211	72,453	+	1,638	Putative
scaffold2628	b2	4,600	6,367	+	1,647	Putative
scaffold2628	b2	25,376	27,100	+	1,362	Putative
scaffold4603	b2	23,182	25,538	+	1,281	Putative
8	c	7,879,444	7,873,491	-	1,011	Putative
scaffold3587	c	24,712	30,665	+	2,229	Putative
8	e	7,893,063	7,889,229	-	2,229	Putative
scaffold3587	e	7,517	11,339	+	1,473	Putative
scaffold3604	f	16,332	8,321	-	1,434	Putative
scaffold3604	f	57,595	51,795	-	2,448	Putative
5	g	31,454,339	31,451,355	-	2,340	Putative
5	g	31,486,606	31,480,702	-	1,254	Putative
5	g	31,507,361	31,505,080	-	1,323	Pseudogene
5	g	31,531,838	31,524,251	-	504	Pseudogene
5	g	31,542,867	31,539,882	-	1,581	Putative
5	g	33,184,300	33,180,563	-	1,254	Putative
scaffold1331	g	68,819	64,340	-	903	Putative
scaffold1331	g	119,024	112,492	-	1,767	Putative
scaffold575	g	87,468	83,011	-	1,719	Putative
scaffold6350	g	6,300	10,193	+	1,767	Putative
scaffold7185	g	2,343	6,197	+	1,767	Putative

Supp 5.4 Gene birth and death rates in the *TPS* gene family across a) multiple taxa and b) the eucalypt lineages



Supp. 5.5 Genomic structure of *TPS* clusters in the *C. citriodora* subsp. *variegata* reference genome assembly (CCV18 v1.1)

This material is unfeasible to present in this thesis due to its format. It can be found alongside the published paper (referred to as Table S4) at: https://www.nature.com/articles/s41437-018-0058-1

Supp. 5.6 *TPS* cluster copy number comparison between *C. citriodora* subsp. *variegata* assemblies CCV18 and CCV54.

	Chr	CCV18 #		CCV54 #
TPS-a	Chr1	1		0
	Chr2	0		2
	Chr3	9		1
	Chr4	3	=	2
		2		0
	Chr5	4	=	1
		1		0
	Chr6	6		7
	Chr7	3		0
		2		0
		1		0
	Chr8	1		3
		1		0
		1		0
	Chr9	1		0
	Chr10	0		1
	Chr11	1		2

	Chr	CCV18 #		CCV54 #
TPS-b1	Chr1	0		1
		8		5
	Chr2	2		0
	Chr4	2		0
	Chr5	3		1
	Chr8	3	=	2
TPS-b2	Chr4	0		2
	Chr6	1		0
	Chr10	3		0
	Chr11	5		1
TPS-c	Chr7	1		0
	Chr8	0		1
TPS-e	Chr6	1		0
	Chr8	0		1
TPS-f	Chr4	3		0
TPS-g	Chr2	2		0
	Chr5	1		0
		3		6

Colour indicates support from CCV54 for the movement of a cluster relative to *E. grandis*, with green showing support for cluster translocation, orange indicating non-support and potential misassembly, and yellow indicating no relevant evidence.

Supp. 5.7 Markers anchoring contigs containing putative translocated *TPS* gene clusters in CCV18

Gene within cluster	Chr	Start position (bp)	Scaffold	Markers from 1CT2-018 linkage map	Markers from 1CT2-050 linkage map	Markers from 1CCV2-054 linkage map
CorciTPS001	1	9,037,469	scaffold1232|size130422	1	2	0
CorciTPS011	4	11,825,354	scaffold1026|size132208	4	2	7
			scaffold1287|size167762	1	0	2
CorciTPS016	5	15,915,747	scaffold1425|size82655	0	1	0
CorciTPS021	6	33,411,585	NA	0	0	0
CorciTPS027	7	1,778,430	scaffold251|size437890	1	0	3
CorciTPS036	9	3,364,546	scaffold214|size477383	0	1	6
			scaffold697|size313272	2	3	2
CorciTPS060	2	2,024,189	scaffold95|size674019	2	3	6
			scaffold24|size1246773	1	5	12
CorciTPS067	8	29,775,806	scaffold727|size227896	3	3	6
			scaffold16|size1201392	5	6	10
CorciTPS079	10	13,586,330	NA	0	0	0
CorciTPS088	7	8,874,051	scaffold627|size532740	5	1	0
			scaffold82|size701448	12	2	9
CorciTPS094	2	19,644,329	scaffold95|size674019	2	3	6
			scaffold24|size1246773	1	5	12

Yellow indicates a scaffold anchored by no markers, green indicates two overlapping scaffolds

Supp 5.8 Gene conversion events in the *TPS-b2* subfamily of *Corymbia citriodora* subsp. *variegata*

Paralog 1	Paralog 2	Corrected p	Start (bp)	End (bp)	Total fragment length	Chr
CorciTPS079	CorciTPS081	$8.69e^{-07}$	1,346	3,096	1,750	10
CorciTPS082	CorciTPS086	$8.47e^{-07}$	1,249	3,175	1,926	11

Gene conversion events between genes hypothesized to be in arrangements lending themselves to concerted evolution were detected using RDP (Martin et al. 2015). The truncated alignment used to construct the phylogeny was recreated using the DNA sequences of the genes in question, which were imported into RDP to perform a full exploratory scan using the default options (see the RDP manual for details). Only one *TPS-b2* clade in CCV was examined, as other clades in CCV and *E. grandis* that seemed subject to concerted evolution had less than three genes present, which is the minimum requirement for this analysis.

Chapter 6 - General discussion

Broad genomic conservation between the eucalypt genera, consistent with other forest trees

The studies presented in this thesis contributes to a view of broad-scale conservation in genomic features between the eucalypt genera *Eucalyptus* and *Corymbia*. This is consistent with past comparative studies investigating population genomic parameters (Hudson et al. 2015) and genome features (Gion et al. 2016) among *Eucalyptus* species*. Eucalyptus* and *Corymbia*, previously considered part of the same genus (Hill and Johnson 1995), show broad conservation at the genome level, with similar chromosomal structure and number of genes in the *TPS* gene family. Comparison with past findings suggests that the scale of differences we observe between *Eucalyptus* and *Corymbia* may be representative of the general conservation of genomic architecture in woody angiosperms. For instance, comparative mapping in *Castanea* (chestnut) and *Quercus* (oak), which diverged approximately 70 million years ago (MYA), revealed conserved chromosome number and high sequence collinearity (Bodénès et al. 2012). Similarly, comparison of genome assemblies of *Salix* (willow) and *Populus* (poplar), which diverged approximately 52 MYA (Tuskan et al. 2006; Dai et al. 2014), revealed high levels of synteny and collinearity between these two genera (Hou et al. 2016), similar to our observations in *Eucalyptus* and *Corymbia.* However, there is one instance of major inter-chromosomal rearrangement, with willow chromosome 1 formed via conjunction of poplar chromosome 16 and part of poplar chromosome 1. No inter-chromosomal translocations were detected in our comparison between the eucalypt chromosomes (although evidence for translocations is seen at the gene level in the *TPS* family). Our comparison was limited in resolution due to comparison of a linkage map and genome assembly, but the recent assembly of the *Corymbia* reference genome has permitted a higher resolution study of synteny and collinearity, which is currently underway (Healey et al. 2017).

Given the long generational time of trees and its implied effect on the rate of molecular evolution, the lack of differentiation described above is not unexpected. Short lived plants generally diverge at a much faster rate, with widespread inter-chromosomal changes reported between many species including *Arabidopsis, Sorghum, Zea, Brassica* and *Fragaria* (Blanc et al. 2000; Swigonova et al. 2004; Parkin et al. 2005; Tennessen et al. 2014). However, the differences in the rate of molecular evolution (of both nuclear sequence and gross chromosomal structure) between plants of different growth habit are more than would be expected from generation time alone (Petit and Hampe 2006), which contributes

to the hypothesis of overall slower evolution in trees (Axelrod 1952; Laroche et al. 1997). An interesting parallel to examine are genomic features of long lived angiosperms and gymnosperms. The gymnosperms (including the four lineages of cycads, *Ginko biloba*, conifers and gnetophytes) are often much longer lived than angiosperms (Bond 1989), which suggests that their rate of molecular evolution could be even slower. Indeed, a comparison of 66 species split between gymnosperms and angiosperms found that silent site divergence (synonymous DNA substitutions that do not change the amino acid sequence of a protein) was on average seven times higher in the angiosperms, pointing to a much higher rate of molecular evolution in angiosperms (De La Torre et al. 2017). This difference persists when considering only woody angiosperms, with poplar estimated to have a four times higher rate of silent site divergence than the gymnosperms *Picea* (spruce) and *Pinus* (pine) (Buschiazzo et al. 2012). This relationship is interesting when considering the often massive genome size of these gymnosperm genera. For example, *C. citriodora* is estimated to have a genome size of approximately 494 MB (Carvalho et al. 2017), almost two orders of magnitude smaller than the 20 - 30 GB of spruce (Nystedt et al. 2013). As no whole genome duplications have been detected in gymnosperms yet (Amanda et al. 2014), one explanation for their bloated genomes is that gymnosperms have been less efficient at preventing the proliferation of transposable elements, leading to the vast non-coding regions of these genomes (Amanda et al. 2014). Angiosperms on the other hand were able to go through 'genome downsizing' by limiting the action of these elements. An interesting physiological consequence of this is a reduction in cell size for the angiosperms, which permits the allocation of a greater density of veins and stomata on leaves and higher photosynthetic efficiency (Simonin and Roddy 2018), and is thought to have contributed to the dominance of the angiosperm lineages.

Fine-scale genomic and genetic differences between eucalypt genera

The comparative analysis based on our high density linkage maps revealed several large intra-chromosomal rearrangements between *Eucalyptus* and *Corymbia,* with most differences verified in *E. grandis, E. globulus, C. citriodora* subsp. *variegata* and *C. torelliana.* Previous comparisons of genome structure within the eucalypts have been limited to more closely related *Eucalyptus* species (Hudson et al. 2012b; Bartholomé et al. 2015), revealing a high degree of collinearity. This comparison of more divergent eucalypts demonstrates the degree of differentiation one can expect in trees given a divergence time of approximately 52 MYA (Thornhill et al. 2015). The position of these specific rearrangements will be relevant for further comparative studies, especially those examining

the position and order of genomic features between organisms, and the genomic basis of speciation.

An example of such an analysis is presented in the annotation of the *terpene synthase* (*TPS*) gene family in *Corymbia*, and the comparison with that of *E. grandis* and *E. globulus*. While similar in overall number and proportional representation of subfamilies (Külheim et al. 2015), *C. citriodora* subsp. *variegata* and *E. grandis* differ significantly in the specific arrangement of these genes. We present evidence for both differential expansion and contraction of *TPS* gene arrays and instances of array translocation and loss, mostly confined to those subfamilies involved in secondary metabolite synthesis. Given that large duplicated gene families such as these are often thought to confer the potential for adaptation (Hanada et al. 2008), the difference in gene numbers observed may be reflective of the different selective pressures experienced by these eucalypt lineages. For instance, the *TPS-b2* family is associated with the synthesis of isoprene, a compound known to play a role in responses to heat stress (Peñuelas et al. 2005). One specific clade of *TPS-b2* genes included five genes in *C. citriodora* subsp. *variegata*, as opposed to two in both *Eucalyptus* species, which given the hotter environments that *Corymbia* is found in may be indicative of environmental adaptation.

Another such examination of genomic features is presented in our analysis of the QTL underlying resistance to various pathogens in *Eucalyptus* and *Corymbia*. We found several QTL for resistance to *Austropuccinia psidii*, an exotic pathogen of global significance, in both *E. globulus* and *Corymbia*. We also examined resistance to several native pathogen genera, including *Teratosphaeria* and *Quambalaria*, the most significant pathogens in eucalypt plantations in temperate and sub-tropical regions, respectively. We present evidence for different QTL underlying separate resistance mechanisms, and the independence of many of these QTL, both within and between genera. For instance, of the 20 QTL discovered in *Corymbia,* there was only a single occurrence of co-location between QTL for resistance to *A. psidii* and *Q. pitereka*. This independence was not limited to the comparison of a native and exotic pathogen, as no co-located QTL were found for resistance to different strains of the same *Q. pitereka* pathogen, suggesting highly specific resistance responses. In contrast, comparison of QTL for disease resistance in *Corymbia* and *Eucalyptus* (including those from independent studies) revealed several notable co-locations between QTL for different pathogens in different host species, suggesting either conserved generalised disease resistance loci or clustering of loci conferring resistance to different pathogens.

Application of these genomic resources

This thesis has produced several resources that will be important for future studies of eucalypt genetic and genomic architecture. These include three high density genetic linkage maps in *Corymbia* spp., which were used to examine differences in chromosomal organisation and the genetic architecture of disease resistance between *Eucalyptus* and *Corymbia*. These linkage maps will provide the foundation for future QTL studies (since the F_1 trees have been planted in field trials (DJ Lee, pers. comm.)) to examine the genetic architecture underlying variation in additional traits of interest in *Corymbia* including developmental traits, growth and wood quality, and will enable comparison with QTL previously identified for these traits in *Eucalyptus* (Freeman et al. 2013; Hudson et al. 2014). The QTL discovered in these *Corymbia* pedigrees adds to the growing literature regarding the genetic architecture underlying variation in disease resistance. These will be useful to provide positional validation for other studies aimed at dissecting the genetic architecture underlying variation in disease resistance, using QTL analysis, expression and association genetic techniques in eucalypts and other taxa (Ingvarsson and Street 2011; Thavamanikumar et al. 2014; Hsieh et al. 2017; Tobias et al. 2017) .

The linkage maps created in this study were key for the assembly of a reference genome for *Corymbia*. The leaps in our understanding of the evolution of the Myrtaceae provided by the *E. grandis* genome is testament to the power of such a resource (Strauss and Myburg 2015). The annotation of genes within the *Corymbia* assembly will facilitate better understanding of gene family evolution between species (Carretero-Paulet et al. 2010), and the identification of genes associated with traits from expression studies. The manual annotation of the terpene synthase gene family, will provide a robust framework to identify the molecular basis of variation in terpene compounds in *Corymbia*, which is particularly important given the role of *Corymbia* species in the essential oil industry (Asante et al. 2001). The examination of other gene families in *Corymbia* and their comparison to other taxa will also be of interest, especially those important to adaptation and evolution such as the nucleotide-binding leucine-rich-repeat family (NLR) which is implicated in pathogen resistance and was recently annotated in *E. grandis* (Christie et al. 2016).

Conclusions

This thesis demonstrates the application of different genomic analyses to the study of divergent species. Important resources were developed for studies ranging from the distribution of individual loci to whole genome comparisons. By examining genomic differentiation among eucalypts at these different scales, we have revealed broad

conservation between different eucalypt genera, along with differences in fine scale features that are likely associated with the different evolutionary histories of these eucalypts. In addition, the resources created for these studies will provide an important foundation for additional examination of various genomic features of *Corymbia* as well as future comparative analyses between other genera.

References

Aharoni A, Giri AP, Deuerlein S, Griepink F, de Kogel W-J, Verstappen FWA, Verhoeven HA, Jongsma MA, Schwab W, Bouwmeester HJ (2003) Terpenoid metabolism in wild-type and transgenic *Arabidopsis* plants. The Plant Cell 15:2866-2884

Ahn S, Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes. Proceedings of the National Academy of Sciences 90:7980-7984

Aitken KS, McNeil MD, Berkman PJ, Hermann S, Kilian A, Bundock PC, Li J (2014) Comparative mapping in the Poaceae family reveals translocations in the complex polyploid genome of sugarcane. BMC Plant Biology 14:1-15

Almeida AC, Landsberg JJ, Sands PJ (2004) Parameterisation of 3-PG model for fast-growing *Eucalyptus grandis* plantations. Forest Ecology and Management 193:179-195

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25:3389-3402

Alves AA, Rosado CCG, Faria DA, da Silva Guimaraes LM, Lau D, Brommonschenkel SH, Grattapaglia D, Alfenas AC (2012) Genetic mapping provides evidence for the role of additive and non-additive QTLs in the response of inter-specific hybrids of *Eucalyptus* to *Puccinia psidii* rust infection. Euphytica 183:27-38

Alves MS, Dadalto SP, Gonçalves AB, De Souza GB, Barros VA, Fietto LG (2013) Plant bZIP transcription factors responsive to pathogens: a review. International Journal of Molecular Sciences 14:7815-7828

Amanda R, Birol I, Bousquet J, Ingvarsson PK, Jansson S, Jones SJ, Keeling CI, MacKay J, Nilsson O, Ritland K (2014) Insights into conifer giga-genomes. Plant Physiology 166:1724-1732

Ammitzboll H, Vaillancourt R, Potts BM, Singarasa S, Mani R, Freeman JS (2018) QTL for intumescence severity in *Eucalyptus globulus* and validation of QTL detection based on phenotyping using open-pollinated families of a mapping population. Plant Disease

Ammon DG, Barton AF, Clarke DA, Tjandra J (1985) Rapid and accurate determination of terpenes in the leaves of *Eucalyptus* species. Analyst 110:921-924

Anagnostakis SL (1987) Chestnut blight: the classical problem of an introduced pathogen. Mycologia 79:23-37

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biology 11:R106

Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. Trends in Ecology & Evolution 19:535-544

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796-815

Asante KS, Brophy JJ, Doran JC, Goldsack RJ, Hibbert DB, Larmour JS (2001) A comparative study of the seedling leaf oils of the spotted gums: species of the *Corymbia* (Myrtaceae), section Politaria. Australian Journal of Botany 49:55-66

Ashton D, Sandiford E (1988) Natural hybridisation between *Eucalyptus regnans* F. Muell. and *E. macrorhyncha* F. Muell. in the Cathedral Range, Victoria. Australian Journal of Botany 36:1-22

Aubourg S, Lecharny A, Bohlmann J (2002) Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. Molecular Genetics and Genomics 267:730-745

Axelrod DI (1952) A theory of angiosperm evolution. Evolution 6:29-60

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75

Baldrich P, Campo S, Wu M-T, Liu T-T, Hsing Y-IC, Segundo BS (2015) MicroRNA-mediated regulation of gene expression in the response of rice plants to fungal elicitors. RNA Biology 12:847-863

Balmelli G, Simeto S, Altier N, Marroni V, Diez JJ (2013) Long term losses caused by foliar diseases on growth and survival of *Eucalyptus globulus* in Uruguay. New Forests 44:249-263

Balmelli G, Simeto S, Marroni V, Altier N, Diez JJ (2014) Genetic variation for resistance to *Mycosphaerella* leaf disease and *Eucalyptus* rust on *Eucalyptus globulus* in Uruguay. Australasian Plant Pathology 43:97-107

Barbour RC, Crawford AC, Henson M, Lee DJ, Potts BM, Shepherd M (2008) The risk of pollen-mediated gene flow from exotic *Corymbia* plantations into native *Corymbia* populations in Australia. Forest Ecology and Management 256:1-19

Bartholomé J, Mandrou E, Mabiala A, Jenkins J, Nabihoudine I, Klopp C, Schmutz J, Plomion C, Gion JM (2015) High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. New Phytologist 206:1283-1296

Batish DR, Singh HP, Kohli RK, Kaur S (2008) *Eucalyptus* essential oil as a natural pesticide. Forest Ecology and Management 256:2166-2174

Beenken L (2017) *Austropuccinia*: a new genus name for the myrtle rust *Puccinia psidii* placed within the redefined family Sphaerophragmiaceae (Pucciniales). Phytotaxa 297:53-61

Benjamini Y, Speed TP (2012) Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Research 40:e72-e72

Berlin S, Lagercrantz U, von Arnold S, Öst T, Rönnberg-Wästljung AC (2010) High-density linkage mapping and evolution of paralogs and orthologs in *Salix* and *Populus*. BMC Genomics 11:1-14

Berthon K, Esperon-Rodriguez M, Beaumont LJ, Carnegie AJ, Leishman MR (2018) Assessment and prioritisation of plant species at risk from myrtle rust

(*Austropuccinia psidii*) under current and future climates in Australia. Biological Conservation 218:154-162

Birney E, Durbin R (2000) Using GeneWise in the *Drosophila* annotation experiment. Genome Research 10:547-548

Bisgrove SR, Simonich MT, Smith NM, Sattler A, Innes RW (1994) A disease resistance gene in *Arabidopsis* with specificity for two different pathogen avirulence genes. The Plant Cell 6:927-933

Blanc G, Barakat A, Guyot R, Cooke R, Delseny M (2000) Extensive duplication and reshuffling in the *Arabidopsis* genome. The Plant Cell 12:1093-1101

Bodénès C, Chancerel E, Gailing O, Vendramin GG, Bagnoli F, Durand J, Goicoechea PG, Soliani C, Villani F, Mattioni C, Koelewijn HP, Murat F, Salse J, Roussel G, Boury C, Alberto F, Kremer A, Plomion C (2012) Comparative mapping in the Fagaceae and beyond with EST-SSRs. BMC Plant Biology 12:1-18

Bond W (1989) The tortoise and the hare: ecology of angiosperm dominance and gymnosperm persistence. Biological Journal of the Linnean Society 36:227-249

Booth TH, Pryor LD (1991) Climatic requirements of some commercially important eucalypt species. Forest Ecology and Management 43:47-60

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32:314-331

Bradshaw HD, Stettler RF (1995) Molecular genetics of growth and development in *Populus*. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree. Genetics 139:963-973

Brawner JT, Lee DJ, Hardner CM, Dieters MJ (2011) Relationships between early growth and *Quambalaria* shoot blight tolerance in *Corymbia citriodora* progeny trials established in Queensland, Australia. Tree Genetics & Genomes 7:759-772

Brondani RP, Williams ER, Brondani C, Grattapaglia D (2006) A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. BMC Plant Biology 6:20

Brooker MIH (2000) A new classification of the genus *Eucalyptus* L'Hér. (Myrtaceae). Australian Systematic Botany 13:79-148

Bundock PC, Potts BM, Vaillancourt RE (2008) Detection and stability of quantitative trait loci (QTL) in *Eucalyptus globulus*. Tree Genetics & Genomes 4:85-95

Burdon JJ, Thrall PH, Ericson, Lars (2006) The current and future dynamics of disease in plant communities. Annual Review of Phytopathology 44:19-39

Burgess TI, Wingfield MJ (2016) Pathogens on the move: a 100-year global experiment with planted eucalypts. Bioscience 67:14-25

Busch M, Seuter A, Hain R (2002) Functional analysis of the early steps of carotenoid biosynthesis in tobacco. Plant Physiology 128:439-453

Buschiazzo E, Ritland C, Bohlmann J, Ritland K (2012) Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. BMC Evolutionary Biology 12:8

Bushnell B (2016) BBMap. http://www.sourceforge.net/projects/bbmap/

Butler JB, Freeman JS, Vaillancourt RE, Potts BM, Glen M, Lee DJ, Pegg GS (2016) Evidence for different QTL underlying the immune and hypersensitive responses of *Eucalyptus globulus* to the rust pathogen *Puccinia psidii*. Tree Genetics & Genomes 12:1-13

Butler JB, Freeman JS, Potts BM, Vaillancourt RE, Grattapaglia D, da Silva Junior OB, Simmons B, Schmutz J, Barry KW, Lee DJ, Healey A, Furtado A, Henry RJ, Baten A, King GJ, Shepherd M (2017a) Comparative analysis of the terpene synthase gene family between *Eucalyptus* and *Corymbia*. In: Plant & Animal Genome Conference XXV, San Diego, United States of America.

Butler JB, Vaillancourt RE, Potts BM, Lee DJ, King GJ, Baten A, Shepherd M, Freeman JS (2017b) Comparative genomics of *Eucalyptus* and *Corymbia* reveals low rates of genome structural rearrangement. BMC Genomics 18:397

Byrne M, Murrell JC, Owen JV, Kriedemann P, Williams ER, Moran GF (1997a) Identification and mode of action of quantitative trait loci affecting seedling height and leaf area in *Eucalyptus nitens*. Theoretical and Applied Genetics 94:674-681

Byrne M, Murrell JC, Owen JV, Williams ER, Moran GF (1997b) Mapping of quantitative trait loci influencing frost tolerance in *Eucalyptus nitens*. Theoretical and Applied Genetics 95:975-979

Canhoto C, Graça M (1999) Leaf barriers to fungal colonization and shredders (*Tipula lateralis*) consumption of decomposing *Eucalyptus globulus*. Microbial Ecology 37:163-172

Cannon SB, Mitra A, Baumgarten A, Young ND, May G (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. BMC Plant Biology 4:10

Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell M (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research 18:188-196

Caplan J, Padmanabhan M, Dinesh-Kumar SP (2008) Plant NB-LRR immune receptors: from recognition to transcriptional reprogramming. Cell Host & Microbe 3:126-135

Card DC, Schield DR, Reyes-Velasco J, Fujita MK, Andrew AL, Oyler-McCance SJ, Fike JA, Tomback DF, Ruggiero RP, Castoe TA (2014) Two low coverage bird genomes and a comparison of reference-guided *versus de novo* genome assemblies. PLoS One 9:e106649

Carnegie AJ (2007) Forest health condition in New South Wales, Australia, 1996–2005. II. Fungal damage recorded in eucalypt plantations during forest health surveys and their management. Australasian Plant Pathology 36:225-239

Carnegie AJ, Lidbetter JR, Walker J, Horwood MA, Tesoriero L, Glen M, Priest MJ (2010) *Uredo rangelii*, a taxon in the guava rust complex, newly recorded on Myrtaceae in Australia. Australasian Plant Pathology 39:463-466

Carnegie AJ, Lidbetter JR (2012) Rapidly expanding host range for *Puccinia psidii* sensu lato in Australia. Australasian Plant Pathology 41:13-29

Carnegie AJ (2015) First report of *Puccinia psidii* (myrtle rust) in *Eucalyptus* plantations in Australia. Plant Disease 99:161-161

Carnegie AJ, Kathuria A, Pegg GS, Entwistle P, Nagel M, Giblin FR (2015) Impact of the invasive rust *Puccinia psidii* (myrtle rust) on native Myrtaceae in natural ecosystems in Australia. Biological Invasions:1-18

Carr D, Carr S (1970) Oil glands and ducts in *Eucalyptus* L'Héit. II. Development and structure of oil glands in the embryo. Australian Journal of Botany 18:191-212

Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martínez-García JF, Bilbao-Castro JR, Robertson DL (2010) Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in *Arabidopsis*, poplar, rice, moss, and algae. Plant Physiology 153:1398-1412

Carvalho GMA, Carvalho CR, Soares FAF (2017) Flow cytometry and cytogenetic tools in eucalypts: genome size variation × karyotype stability. Tree Genetics & Genomes 13:106

Cesari S, Thilliez G, Ribot C, Chalvon V, Michel C, Jauneau A, Rivas S, Alaux L, Kanzaki H, Okuyama Y, Morel J-B, Fournier E, Tharreau D, Terauchi R, Kroj T (2013) The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding. The Plant Cell 25:1463-1481

Cesari S (2017) Multiple strategies for pathogen perception by plant immune receptors. New Phytologist

Chaw S-M, Chang C-C, Chen H-L, Li W-H (2004) Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. Journal of Molecular Evolution 58:424-441

Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. The Plant Journal 66:212-229

Chen J-M, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP (2007) Gene conversion: mechanisms, evolution and human disease. Nature Reviews: Genetics 8:762-775

Chen S-C, Cannon CH, Kua C-S, Liu J-J, Galbraith DW (2014) Genome size variation in the Fagaceae and its implications for trees. Tree Genetics & Genomes 10:977-988

Chittenden L, Schertz K, Lin Y, Wing R, Paterson A (1994) A detailed RFLP map of *Sorghum bicolor* x *S. propinquum*, suitable for high-density mapping, suggests ancestral duplication of *Sorghum* chromosomes or chromosomal segments. Theoretical and Applied Genetics 87:925-933

Christie N, Tobias PA, Naidoo S, Külheim C (2016) The *Eucalyptus grandis* NBS-LRR gene family: physical clustering and expression hotspots. Frontiers in Plant Science 6:1238

Chu HY, Wegel E, Osbourn A (2011) From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. The Plant Journal 66:66-79

Chu Z, Ouyang Y, Zhang J, Yang H, Wang S (2004) Genome-wide analysis of defense-responsive genes in bacterial blight resistance of rice mediated by the recessive R gene xa13. Molecular Genetics and Genomics 271:111-120

Church GM, Gilbert W (1984) Genomic sequencing. Proceedings of the National Academy of Sciences 81:1991-1995

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963-971

Collard B, Jahufer M, Brouwer J, Pang E (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169-196

Collard B, Mace E, McPhail M, Wenzl P, Cakir M, Fox G, Poulsen D, Jordan D (2009) How accurate are the marker orders in crop linkage maps generated from large marker datasets? Crop and Pasture Science 60:362-372

Coutinho T, Wingfield M, Alfenas A, Crous P (1998) *Eucalyptus* rust: a disease with the potential for serious international implications. Plant Disease 82:819-825

Crabill E, Joe A, Block A, van Rooyen JM, Alfano JR (2010) Plant immunity directly or indirectly restricts the injection of type III effectors by the *Pseudomonas syringae* type III secretion system. Plant Physiology 154:233-244

Crisp MD, Burrows GE, Cook LG, Thornhill AH, Bowman DMJS (2011) Flammable biomes dominated by eucalypts originated at the Cretaceous-Palaeogene boundary. Nature Communications 2:193

Cui H, Tsuda K, Parker JE (2015) Effector-Triggered Immunity: From pathogen perception to robust defense. Annual Review of Plant Biology 66:487-511

Cui L-G, Shan J-X, Shi M, Gao J-P, Lin H-X (2014) The miR156-SPL9-DFR pathway coordinates the relationship between development and abiotic stress tolerance in plants. The Plant Journal 80:1108-1117

Dai X, Hu Q, Cai Q, Feng K, Ye N, Tuskan GA, Milne R, Chen Y, Wan Z, Wang Z, Luo W, Wang K, Wan D, Wang M, Wang J, Liu J, Yin T (2014) The willow genome and divergent evolution from poplar after the common genome duplication. Cell Research 24:1274

Dandekar T, Snel B, Huynen M, Bork P (1998) Conservation of gene order: a fingerprint of proteins that physically interact. Trends in Biochemical Sciences 23:324-328

Dangl JL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. Nature 411:826-833

De La Torre AR, Li Z, Van de Peer Y, Ingvarsson PK (2017) Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. Molecular Biology and Evolution 34:1363-1377

del Moral R, Muller CH (1970) The allelopathic effects of *Eucalyptus camaldulensis*. The American Midland Naturalist 83:254-282

Demuth JP, Hahn MW (2009) The life and death of gene families. Bioessays 31:29-39

DeYoung BJ, Qi D, Kim S-H, Burke TP, Innes RW (2012) Activation of a plant nucleotide binding-leucine rich repeat disease resistance protein by a modified self protein. Cellular Microbiology 14:1071-1084

Dickinson GR, Lee DJ, Huth JR (2004) Early plantation growth and tolerance to ramularia shoot blight of provenances of three spotted gum taxa on a range of sites in Queensland. Australian Forestry 67:122-130

Dickinson GR, Wallace HM, Lee DJ (2010) Controlled pollination methods for creating *Corymbia* hybrids. Silvae Genetica 59:233

Dickinson GR, Lee DJ, Wallace HM (2012) The influence of pre- and post-zygotic barriers on interspecific *Corymbia* hybridization. Annals of Botany 109:1215-1226

Dillon SK, Brawner JT, Meder R, Lee DJ, Southerton SG (2012) Association genetics in *Corymbia citriodora* subsp. *variegata* identifies single nucleotide polymorphisms affecting wood growth and cellulosic pulp yield. New Phytologist 195:596-608

Dodds PN, Lawrence GJ, Catanzariti A-M, Teh T, Wang C-IA, Ayliffe MA, Kobe B, Ellis JG (2006) Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. Proceedings of the National Academy of Sciences 103:8888-8893

Dodds PN, Rathjen JP (2010) Plant immunity: towards an integrated view of plant–pathogen interactions. Nature Reviews Genetics 11:539-548

Doughty RW (2000) The *Eucalyptus*: a natural and commercial history of the gum tree. Johns Hopkins University Press, Baltimore and London

Douglas MH, van Klink JW, Smallfield BM, Perry NB, Anderson RE, Johnstone P, Weavers RT (2004) Essential oils from New Zealand manuka: triketone and other chemotypes of *Leptospermum scoparium*. Phytochemistry 65:1255-1264

Dowkiw A, Voisin E, Bastien C (2010) Potential of Eurasian poplar rust to overcome a major quantitative resistance factor. Plant Pathology 59:523-534

DPIPWE (2018) Plant species affected by myrtle rust in Tasmania. http://dpipwe.tas.gov.au/biosecurity-tasmania/plant-biosecurity/pests-and-diseases/myrtle-rust/plant-species-affected-by-myrtle-rust-in-tasmania. Accessed 28/04/2018 2018

Drosophila 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450:203-218

Dutkowski G, Potts B (1999) Geographic patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification. Australian Journal of Botany 47:237-263

Dvořák J, Zhang HB (1992) Application of molecular tools for study of the phylogeny of diploid and polyploid taxa in Triticeae. Hereditas 116:37-42

Ee S-F, Mohamed-Hussein Z-A, Othman R, Shaharuddin NA, Ismail I, Zainal Z (2014) Functional characterization of sesquiterpene synthase from *Polygonum minus*. The Scientific World Journal 2014:11

Eitas TK, Dangl JL (2010) NB-LRR proteins: pairs, pieces, perception, partners, and pathways. Current Opinion in Plant Biology 13:472-477

Ellis J, Jones D (1998) Structure and function of proteins controlling strain-specific pathogen resistance in plants. Current Opinion in Plant Biology 1:288-293

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57

Ennos RA (2014) Resilience of forests to pathogens: an evolutionary ecology perspective. Forestry: An International Journal of Forest Research 88:41-52

Eulgem T, Somssich IE (2007) Networks of WRKY transcription factors in defense signaling. Current Opinion in Plant Biology 10:366-371

Fawal N, Li Q, Mathé C, Dunand C (2014) Automatic multigenic family annotation: risks and solutions. Trends in Genetics 30:323-325

Fenster CB, Galloway LF (2000) Population differentiation in an annual legume: genetic architecture. Evolution 54:1157-1172

Ferreira F (1983) *Eucalyptus* rust. Revista Arvore 7:91-109

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. Nature Reviews: Genetics 3:329-341

Field B, Osbourn AE (2008) Metabolic diversification - independent assembly of operon-like gene clusters in different plants. Science 320:543-547

Field B, Fiston-Lavier A-S, Kemen A, Geisler K, Quesneville H, Osbourn AE (2011) Formation of plant metabolic gene clusters within dynamic chromosomal regions. Proceedings of the National Academy of Sciences 108:16116-16121

Fisher RA (1930) The genetical theory of natural selection. University Press, Oxford

Fishman L, Willis JH, Wu CA, Lee YW (2014) Comparative linkage maps suggest that fission, not polyploidy, underlies near-doubling of chromosome number within monkeyflowers (*Mimulus*; Phrymaceae). Heredity 112:562-568

Flagel LE, Wendel JF (2009) Gene duplication and evolutionary novelty in plants. New Phytologist 183:557-564

Flor H (1942) Inheritance of pathogenicity in *Melampsora lini*. Phytopathology 32:e69

Florea L, Di Francesco V, Miller J, Turner R, Yao A, Harris M, Walenz B, Mobarry C, Merkulov GV, Charlab R (2005) Gene and alternative splicing annotation with AIR. Genome Research 15:54-66

Fray RG, Wallace A, Fraser PD, Valero D, Hedden P, Bramley PM, Grierson D (1995) Constitutive expression of a fruit phytoene synthase gene in transgenic tomatoes causes dwarfism by redirecting metabolites from the gibberellin pathway. The Plant Journal 8:693-701

Freeling M (2009) Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annual Review of Plant Biology 60:433-453

Freeman JS, O'Reilly-Wapstra JM, Vaillancourt RE, Wiggins N, Potts BM (2008a) Quantitative trait loci for key defensive compounds affecting herbivory of eucalypts in Australia. New Phytologist 178:846-851

Freeman JS, Potts BM, Vaillancourt RE (2008b) Few Mendelian genes underlie the quantitative response of a forest tree, *Eucalyptus globulus*, to a natural fungal epidemic. Genetics 178:563-571

Freeman JS, Whittock SP, Potts BM, Vaillancourt RE (2009) QTL influencing growth and wood properties in *Eucalyptus globulus*. Tree Genetics & Genomes 5:713-722

Freeman JS, Potts BM, Downes GM, Pilbeam D, Thavamanikumar S, Vaillancourt RE (2013) Stability of quantitative trait loci for growth and wood properties across multiple pedigrees and environments in *Eucalyptus globulus*. New Phytologist 198:1121-1134

Freeman JS (2014) Molecular linkage maps of *Eucalyptus*: strategies, resources and achievements. In: Genetics, Genomics and Breeding of Eucalypts. CRC Press, Boca Raton Florida, p 58

Gao Y, Honzatko RB, Peters RJ (2012) Terpenoid synthase structures: a so far incomplete view of complex catalysis. Natural Product Reports 29:1153-1175

Gavran M (2014) Australian Plantation Statistics 2014 Update. Australian Bureau of Agricultural and Resource Economics and Sciences, Canberra

Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139:907-920

Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, O'shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. Nature 425:737

Giblin F, Carnegie A (2014) *Puccinia psidii* (Myrtle rust) - Australian host list. http://www.anpc.asn.au/myrtle-rust. Accessed 10 May 2018

Gijzen M (2009) Runaway repeats force expansion of the *Phytophthora infestans* genome. Genome Biology 10:241

Gion J-M, Hudson CJ, Lesur I, Vaillancourt RE, Potts BM, Freeman JS (2016) Genome-wide variation in recombination rate in *Eucalyptus*. BMC Genomics 17:590

Glen M, Alfenas A, Zauza E, Wingfield M, Mohammed C (2007) *Puccinia psidii*: a threat to the Australian environment and economy—a review. Australasian Plant Pathology 36:1-16

Goodger JQ, Heskes AM, Woodrow IE (2013) Contrasting ontogenetic trajectories for phenolic and terpenoid defences in *Eucalyptus froggattii*. Annals of Botany 112:651-659

Gosney BJ, Potts BM, O'Reilly-Wapstra JM, Vaillancourt RE, Fitzgerald H, Davies NW, Freeman JS (2016) Genetic control of cuticular wax compounds in *Eucalyptus globulus*. New Phytologist 209:202-215

Gosney BJ (2017) Community genetics of eucalypts: Provenance effects on canopy communities, potential drivers and underlying QTL. PhD thesis, University of Tasmania

Graça RN, Aun CP, Guimarães LM, Rodrigues BV, Zauza EA, Alfenas AC (2011) A new race of *Puccinia psidii* defeats rust resistance in eucalypt. Australasian Plant Pathology 40:442-447

Graça RN, Ross-Davis AL, Klopfenstein NB, Kim M-S, Peever TL, Cannon PG, Aun CP, Mizubuti ESG, Alfenas AC (2013) Rust disease of eucalypts, caused by *Puccinia psidii*, did not originate via host jump from guava in Brazil. Molecular Ecology 22:6033-6047

Grattapaglia D, Bradshaw Jr HD (1994) Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. Canadian Journal of Forest Research 24:1074-1078

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137:1121-1137

Grattapaglia D, Bertolucci FL, Sederoff RR (1995) Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross strategy and RAPD markers. Theoretical and Applied Genetics 90:933-947

Grattapaglia D, Kirst M (2008) *Eucalyptus* applied genomics: from gene sequences to breeding tools. New Phytologist 179:911-929

Grattapaglia D, Resende MD (2011) Genomic selection in forest tree breeding. Tree Genetics & Genomes 7:241-255

Grattapaglia D, Vaillancourt RE, Shepherd M, Thumma BR, Foley W, Külheim C, Potts BM, Myburg AA (2012) Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. Tree Genetics & Genomes 8:463-508

Graves S, Piepho H, Selzer L, Dorai-Raj S (2012) multcompView: Visualizations of Paired Comparisons. R package version 01-5, URL http://CRANR-projectorg/package=multcompView

Griffin A, Burgess I, Wolf L (1988) Patterns of natural and manipulated hybridisation in the genus *Eucalyptus* L'hérit. - a review. Australian Journal of Botany 36:41-66

Groenen MAM, Wahlberg P, Foglio M, Cheng HH, Megens H-J, Crooijmans RPMA, Besnier F, Lathrop M, Muir WM, Wong GK-S, Gut I, Andersson L (2009) A high-density SNP-

based linkage map of the chicken genome reveals sequence features correlated with recombination rate. Genome Research 19:510-519

Gururani MA, Venkatesh J, Upadhyaya CP, Nookaraju A, Pandey SK, Park SW (2012) Plant disease resistance genes: current status and future directions. Physiological and Molecular Plant Pathology 78:51-65

Gustafson JP, Ma X-F, Korzun V, Snape JW (2009) A consensus map of rye integrating mapping data from five mapping populations. Theoretical and Applied Genetics 118:793-800

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A (2013) *De novo* transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nature Protocols 8:10.1038/nprot.2013.1084

Hackett CA, Broadfoot LB (2003) Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. Heredity 90:33-38

Hadley W (2009) ggplot2: Elegant graphics for Data Analysis. Springer, New York

Hamilton JP, Robin Buell C (2012) Advances in plant genome sequencing. The Plant Journal 70:177-190

Hamilton M, Joyce K, Williams D, Dutkowski G, Potts B (2008) Achievements in forest tree improvement in Australia and New Zealand 9. Genetic improvement of *Eucalyptus nitens* in Australia. Australian Forestry 71:82-93

Hamilton M, Williams D, Tilyard P, Pinkard E, Wardlaw T, Glen M, Vaillancourt R, Potts B (2013) A latitudinal cline in disease resistance of a host tree. Heredity 110:372-379

Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu S-H (2008) Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. Plant Physiology 148:993-1003

Hanley SJ, Mallott MD, Karp A (2006) Alignment of a *Salix* linkage map to the *Populus* genomic sequence reveals macrosynteny between willow and poplar genomes. Tree Genetics & Genomes 3:35-48

Hansen NL, Heskes AM, Hamberger B, Olsen CE, Hallström BM, Andersen-Ranberg J, Hamberger B (2017) The terpene synthase gene family in *Tripterygium wilfordii* harbors a labdane-type diterpene synthase among the monoterpene synthase TPS-b subfamily. The Plant Journal 89:429-441

Hansen TF (2006) The evolution of genetic architecture. Annual Review of Ecology, Evolution, and Systematics 37:123-157

Hayashi K-i, Kawaide H, Notomi M, Sakigi Y, Matsuo A, Nozaki H (2006) Identification and functional analysis of bifunctional ent-kaurene synthase from the moss *Physcomitrella patens*. FEBS Letters 580:6175-6181

Hayes BJ, Bowman PJ, Chamberlain A, Goddard M (2009) Invited review: Genomic selection in dairy cattle: Progress and challenges. Journal of Dairy Science 92:433-443

Healey A, Shepherd M, Baten A, King GJ, Lee DJ, Furtado A, Vaillancourt RE, Butler JB, Freeman JS, Potts BM, Grattapaglia D, Junior OBdS, Barry KW, Schmutz J, Simmons B, Henry RJ (2017) Sequencing the branches of the eucalypt tree: comparison between *Eucalyptus* and *Corymbia* genomes. In: Plant & Animal Genome Conference XXV, San Diego, United States of America.

Heath MC (2000a) Hypersensitive response-related death. In: Lam E, Fukuda H, Greenberg J (eds) Programmed Cell Death in Higher Plants. Springer, Netherlands, pp 77-90

Heath MC (2000b) Nonhost resistance and nonspecific plant defenses. Current Opinion in Plant Biology 3:315-319

Heiling S, Schuman MC, Schoettner M, Mukerjee P, Berger B, Schneider B, Jassbi AR, Baldwin IT (2010) Jasmonate and ppHsystemin regulate key malonylation steps in the biosynthesis of 17-hydroxygeranyllinalool diterpene glycosides, an abundant and effective direct defense against herbivores in *Nicotiana attenuata*. The Plant Cell 22:273-292

Henery ML, Moran GF, Wallis IR, Foley WJ (2007) Identification of quantitative trait loci influencing foliar concentrations of terpenes and formylated phloroglucinol compounds in *Eucalyptus nitens*. New Phytologist 176:82-95

Heslot N, Rutkoski J, Poland J, Jannink J-L, Sorrells ME (2013) Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. PLoS One 8:e74612

Hill KD, Johnson LA (1995) Systematic studies in the eucalypts 7. A revision of the bloodwoods, genus *Corymbia* (Myrtaceae). Telopea 6:185-504

Hobert O (2008) Gene regulation by transcription factors and microRNAs. Science 319:1785-1786

Hofberger JA, Lyons E, Edger PP, Chris Pires J, Eric Schranz M (2013) Whole genome and tandem duplicate retention facilitated glucosinolate pathway diversification in the mustard family. Genome Biology and Evolution 5:2155-2173

Hofberger JA, Nsibo DL, Govers F, Bouwmeester K, Schranz ME (2015) A complex interplay of tandem- and whole-genome duplication drives expansion of the L-type lectin receptor kinase gene family in the Brassicaceae. Genome Biology and Evolution 7:720-734

Homer LE, Leach DN, Lea D, Slade Lee L, Henry RJ, Baverstock PR (2000) Natural variation in the essential oil content of *Melaleuca alternifolia* Cheel (Myrtaceae). Biochemical Systematics and Ecology 28:367-382

Hörger AC, Ilyas M, Stephan W, Tellier A, van der Hoorn RAL, Rose LE (2012) Balancing selection at the tomato RCR3 guardee gene family maintains variation in strength of pathogen defense. PLoS Genetics 8:e1002813

Hosfield DJ, Zhang Y, Dougan DR, Broun A, Tari LW, Swanson RV, Finn J (2004) Structural basis for bisphosphonate-mediated inhibition of isoprenoid biosynthesis. Journal of Biological Chemistry 279:8526-8529

Hou J, Ye N, Dong Z, Lu M, Li L, Yin T (2016) Major chromosomal rearrangements distinguish willow and poplar after the ancestral “Salicoid” genome duplication. Genome Biology and Evolution 8:1868-1875

Howe D, Costanzo M, Fey P, Gojobori T, Hannick L, Hide W, Hill DP, Kania R, Schaeffer M, St Pierre S (2008) Big data: the future of biocuration. Nature 455:47

Hsieh J-F, Chuah A, Patel HR, Sandhu KS, Foley WJ, Külheim C (2017) Transcriptome profiling of *Melaleuca quinquenervia* challenged by myrtle rust reveals differences in defense responses among resistant individuals. Phytopathology 108:495-509

Huang J, Gu M, Lai Z, Fan B, Shi K, Zhou Y-H, Yu J-Q, Chen Z (2010) Functional analysis of the *Arabidopsis* PAL gene family in plant growth, development, and response to environmental stress. Plant Physiology 153:1526-1538

Hüberli D, Tommerup IC, Dobrowolski MP, Calver MC, St J Hardy GE (2001) Phenotypic variation in a clonal lineage of two *Phytophthora cinnamomi* populations from Western Australia. Mycological Research 105:1053-1064

Hudson CJ, Freeman JS, Kullan AR, Petroli CD, Sansaloni CP, Kilian A, Detering F, Grattapaglia D, Potts BM, Myburg AA (2012a) A reference linkage map for *Eucalyptus*. BMC Genomics 13:240

Hudson CJ, Kullan ARK, Freeman JS, Faria DA, Grattapaglia D, Kilian A, Myburg AA, Potts BM, Vaillancourt RE (2012b) High synteny and colinearity among *Eucalyptus* genomes revealed by high-density comparative genetic mapping. Tree Genetics & Genomes 8:339-352

Hudson CJ, Freeman JS, Jones RC, Potts BM, Wong MM, Weller JL, Hecht VF, Poethig RS, Vaillancourt RE (2014) Genetic control of heterochrony in *Eucalyptus globulus*. G3 4:1235-1245

Hudson CJ, Freeman JS, Myburg AA, Potts BM, Vaillancourt RE (2015) Genomic patterns of species diversity and divergence in *Eucalyptus*. New Phytologist 206:1378-1390

Hurles M (2004) Gene duplication: the genomic trade in spare parts. PLoS Biology 2:e206

Iglesias-Trabado G, Carbaeira-Tenreiro R, Folgueiia-Lozano J (2009) *Eucalyptus universalis:* Global cultivated eucalypt forest map (Version 1.2). In: Eucalyptologics: Information resources on *Eucalyptus* cultivation worldwide. Retrieved from http://www.git-forestry.com (September 16th, 2015).

Ingvarsson PK, Street NR (2011) Association genetics of complex traits in plants. New Phytologist 189:909-922

International Barley Genome Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711-716

Irmisch S, Jiang Y, Chen F, Gershenzon J, Köllner TG (2014) Terpene synthases and their contribution to herbivore-induced volatile emission in western balsam poplar (*Populus trichocarpa*). BMC Plant Biology 14:270

Iskow RC, Gokcumen O, Lee C (2012) Exploring the role of copy number variants in human adaptation. Trends in Genetics 28:245-257

Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology 3:318-356

Jacob F, Vernaldi S, Maekawa T (2013) Evolution and conservation of plant NLR functions. Frontiers in Immunology 4:297

Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Fabbro CD, Alaux M, Gaspero GD, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Clainche IL, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pè ME, Valle G, Morgante M, Caboche M, Adam-Blondon A-F, Weissenbach J, Quétier F, Wincker P (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463-467

Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Briefings in Functional Genomics 9:166-177

Jansen C, Von Wettstein D, Schäfer W, Kogel K-H, Felk A, Maier FJ (2005) Infection patterns in barley and wheat spikes inoculated with wild-type and trichodiene synthase gene disrupted *Fusarium graminearum*. Proceedings of the National Academy of Sciences of the United States of America 102:16892-16897

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447-1455

Jenkins CC, Suberkropp K (1995) The influence of water chemistry on the enzymatic degradation of leaves in streams. Freshwater Biology 33:245-253

Jighly A, Joukhadar R, Alagu M (2015) SimpleMap: a pipeline to streamline high-density linkage map construction. The Plant Genome 8

Johnson G, Burdon R (1990) Family-site interaction in *Pinus radiata*: implications for progeny testing strategy and regionalised breeding in New Zealand. Silvae Genetica 39:55-62

Johnson IG, Carnegie AJ, Henson M (2009) Growth, form and Quambalaria shoot blight tolerance of spotted gum in north-eastern New South Wales, Australia. Silvae Genetica 58:180

Jones JD, Dangl JL (2006) The plant immune system. Nature 444:323-329

Junghans D, Alfenas A, Brommonschenkel S, Oda S, Mello E, Grattapaglia D (2003a) Resistance to rust (*Puccinia psidii* Winter) in *Eucalyptus*: mode of inheritance and mapping of a major gene with RAPD markers. Theoretical and Applied Genetics 108:175-180

Junghans DT, Alfenas AC, Maffia LA (2003b) Escala de notas para quantificação da ferrugem em *Eucalyptus*. Fitopatologia Brasileira 28:184-188

Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H (2014) Efficient *de novo* assembly of highly heterozygous

genomes from whole-genome shotgun short reads. Genome Research 24:1384-1395

Karnosky DF (1979) Dutch elm disease: a review of the history, environmental implications, control, and research needs. Environmental Conservation 6:311-322

Kashi Y, Hallerman E, Soller M (1990) Marker-assisted selection of candidate bulls for progeny testing programmes. Animal Production 51:63-74

Kawakami T, Smeds L, Backström N, Husby A, Qvarnström A, Mugal CF, Olason P, Ellegren H (2014) A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Molecular Ecology 23:4035-4058

Kawanishi T, Uematsu S, Kakishima M, Kagiwada S, Hamamoto H, Horie H, Namba S (2009) First report of rust disease on ohia and the causal fungus, *Puccinia psidii*, in Japan. Journal of General Plant Pathology 75:428-431

Keith R, Mitchell-Olds T (2013) Genetic variation for resistance to herbivores and plant pathogens: hypotheses, mechanisms and evolutionary implications. Plant Pathology 62:122-132

Keszei A, Brubaker CL, Foley WJ (2008) A molecular perspective on terpene variation in Australian Myrtaceae. Australian Journal of Botany 56:197-213

Keszei A, Brubaker CL, Carter R, Köllner T, Degenhardt J, Foley WJ (2010a) Functional and evolutionary relationships between terpene synthases from Australian Myrtaceae. Phytochemistry 71:844-852

Keszei A, Hassan Y, Foley WJ (2010b) A biochemical interpretation of terpene chemotypes in *Melaleuca alternifolia*. Journal of Chemical Ecology 36:652-661

Kim YJ, Lin N-C, Martin GB (2002) Two distinct *Pseudomonas* effector proteins interact with the Pto kinase and activate plant immunity. Cell 109:589-598

Kinloch BB, Dupper GE (2002) Genetic specificity in the white pine-blister rust pathosystem. Phytopathology 92:278-280

Klement Z, Bozsó Z, Kecskés ML, Besenyei E, Arnold C, Ott PG (2003) Local early induced resistance of plants as the first line of defence against bacteria. Pest Management Science 59:465-474

Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T (2001) Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate–dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. The Plant Cell 13:681-694

Kole C, Williams PH, Rimmer SR, Osborn TC (2002) Linkage mapping of genes controlling resistance to white rust (*Albugo candida*) in *Brassica rapa* (syn. *campestris*) and comparative mapping to *Brassica napus* and *Arabidopsis thaliana*. Genome 45:22-27

Kondrashov FA (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. Proceedings of the Royal Society B: Biological Sciences 279:5048-5057

Koonin EV, Mushegian AR, Rudd KE (1996) Sequencing and analysis of bacterial genomes. Current Biology 6:404-416

Koonin EV, Wolf YI (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Research 36:6688-6719

Koonin EV (2009) Evolution of genome architecture. The International Journal of Biochemistry & Cell Biology 41:298-306

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29

Kosambi DD (1943) The estimation of map distances from recombination values. Annals of Eugenics 12:172-175

Kriticos DJ, Morin L, Leriche A, Anderson RC, Caley P (2013) Combining a climatic niche model of an invasive fungus with its host species distributions to identify risks to natural assets: *Puccinia psidii* sensu lato in Australia. PLoS ONE 8:e64479

Kukekova AV, Trut LN, Oskina IN, Johnson JL, Temnykh SV, Kharlamova AV, Shepeleva DV, Gulievich RG, Shikhevich SG, Graphodatsky AS, Aguirre GD, Acland GM (2007) A meiotic linkage map of the silver fox, aligned and compared to the canine genome. Genome Research 17:387-399

Külheim C, Padovan A, Hefer C, Krause ST, Köllner TG, Myburg AA, Degenhardt J, Foley WJ (2015) The *Eucalyptus* terpene synthase gene family. BMC Genomics 16:1-18

Kullan ARK, van Dyk MM, Jones N, Kanzler A, Bayley A, Myburg AA (2011) High-density genetic linkage maps with over 2,400 sequence-anchored DArT markers for genetic dissection in an F2 pseudo-backcross of *Eucalyptus grandis* × *E. urophylla*. Tree Genetics & Genomes 8:163-175

Kushalappa AC, Yogendra KN, Karre S (2016) Plant innate immune response: qualitative and quantitative resistance. Critical Reviews in Plant Sciences 35:38-55

Labate CA, Assis TF, Oda S, Mello EJ, González ER, Zauza EAV, Mori ES, Moraes MLT, Cid LPB, Alfenas AC (2009) *Eucalyptus*. Compendium of Transgenic Crop Plants 9:35-108

Ladiges PY, Udovicic F, Nelson G (2003) Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. Journal of Biogeography 30:989-998

Lanfear R, Ho SYW, Jonathan Davies T, Moles AT, Aarssen L, Swenson NG, Warman L, Zanne AE, Allen AP (2013) Taller plants have lower rates of molecular evolution. Nat Commun 4:1879

Larcombe MJ, Holland B, Steane DA, Jones RC, Nicolle D, Vaillancourt RE, Potts BM (2015) Patterns of reproductive isolation in *Eucalyptus* - a phylogenetic perspective. Molecular Biology and Evolution 32:1833-1846

Laroche J, Li P, Maggia L, Bousquet J (1997) Molecular evolution of angiosperm mitochondrial introns and exons. Proceedings of the National Academy of Sciences 94:5722-5727

Laurie CC, Chasalow SD, LeDeaux JR, McCarroll R, Bush D, Hauge B, Lai C, Clark D, Rocheford TR, Dudley JW (2004) The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics 168:2141-2155

Lawler IR, Foley WJ, Eschler BM, Pass DM, Handasyde K (1998) Intraspecific variation in *Eucalyptus* secondary metabolites determines food intake by folivorous marsupials. Oecologia 116:160-169

Lawler IR, Stapley J, Foley WJ, Eschler BM (1999) Ecological example of conditioned flavor aversion in plant–herbivore interactions: effect of terpenes of *Eucalyptus l*eaves on feeding by common ringtail and brushtail possums. Journal of Chemical Ecology 25:401-415

Lawson SA, McDonald JM, Pegg GS (2008) Forest health surveillance methodology in hardwood plantations in Queensland, Australia. Australian Forestry 71:177-181

Lee DJ (2007) Achievements in forest tree genetic improvement in Australia and New Zealand 2: Development of *Corymbia* species and hybrids for plantations in eastern Australia. Australian Forestry 70:11-16

Lee DJ, Huth JR, Brawner JT, Dickinson GR (2009) Comparative performance of *Corymbia* hybrids and parental species in subtropical Queensland and implications for breeding and deployment. Silvae Genetica 58:205

Lee H-A, Yeom S-I (2015) Plant NB-LRR proteins: tightly regulated sensors in a complex manner. Briefings in Functional Genomics 14:233-242

Lee JM, Sonnhammer ELL (2003) Genomic gene clustering analysis of pathways in eukaryotes. Genome Research 13:875-882

Lefort V, Longueville J-E, Gascuel O (2017) SMS: Smart Model Selection in PhyML. Molecular Biology and Evolution 34:2422-2424

Leister D (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends in Genetics 20:116-122

Li F, Zhou C, Weng Q, Li M, Yu X, Guo Y, Wang Y, Zhang X, Gan S (2015) Comparative genomics analyses reveal extensive chromosome colinearity and novel quantitative trait loci in *Eucalyptus*. PLoS ONE 10:e0145144

Li H, Ribaut J-M, Li Z, Wang J (2008) Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. Theoretical and Applied Genetics 116:243-260

Librado P, Vieira FG, Rozas J (2012) BadiRate: estimating family turnover rates by likelihood-based methods. Bioinformatics 28:279-281

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. American Journal of Human Genetics 44:397

Liu C-C, Li C-M, Liu B-G, Ge S-J, Dong X-M, Li W, Zhu H-Y, Wang B-C, Yang C-P (2012) Genome-wide identification and characterization of a dehydrin gene family in Poplar (*Populus trichocarpa*). Plant Molecular Biology Reporter 30:848-859

Liu X, Chen Q, Wang Z, Xie L, Xu Z (2008) Allelopathic effects of essential oil from *Eucalyptus grandis* × *E*. *urophylla* on pathogenic fungi and pest insects. Frontiers of Forestry in China 3:232-236

Liu Y, Schiff M, Dinesh-Kumar S (2004) Involvement of MEK1 MAPKK, NTF6 MAPK, WRKY/MYB transcription factors, COI1 and CTR1 in N-mediated resistance to tobacco mosaic virus. The Plant Journal 38:800-809

Long M, Betran E, Thornton K, Wang W (2003) The origin of new genes: glimpses from the young and old. Nature Reviews Genetics 4:865-875

Lu Y, Hatsugai N, Katagiri F, Ishimaru CA, Glazebrook J (2015) Putative serine protease effectors of *Clavibacter michiganensis* induce a hypersensitive response in the apoplast of *Nicotiana* species. Molecular Plant-Microbe Interactions:MPMI-02-15-0036-R

Luo M-C, You FM, Li P, Wang J-R, Zhu T, Dandekar AM, Leslie CA, Aradhya M, McGuire PE, Dvorak J (2015) Synteny analysis in Rosids with a walnut physical map reveals slow genome evolution in long-lived woody perennials. BMC Genomics 16:1-17

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits vol 1. Sinauer Associates, Inc., Sunderland, MA

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151-1155

Lynch M (2007) The origins of genome architecture. Sinauer Associates, Inc., Sunderland, MA

Lyons E, Freeling M (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. The Plant Journal 53:661-673

Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang H, Wang X, Bowers J, Paterson A, Lisch D, Freeling M (2008) Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with rosids. Plant Physiology 148:1772-1781

Mace ES, Rami J-F, Bouchet S, Klein PE, Klein RR, Kilian A, Wenzl P, Xia L, Halloran K, Jordan DR (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. BMC Plant Biology 9:13

Machado P, Alfenas A, Alfenas R, Mohammed C, Glen M (2015) Microsatellite analysis indicates that *Puccinia psidii* in Australia is mutating but not recombining. Australasian Plant Pathology 44:455-462

Mackay TFC (2001) The genetic architecture of quantitative traits. Annual Review of Genetics 35:303-339

Macphail M, Thornhill AH (2016) How old are the eucalypts? A review of the microfossil and phylogenetic evidence. Australian Journal of Botany 64:579-599

Mamani EM, Bueno NW, Faria DA, Guimarães LM, Lau D, Alfenas AC, Grattapaglia D (2010) Positioning of the major locus for *Puccinia psidii* rust resistance (*Ppr1*) on the

Eucalyptus reference map and its validation across unrelated pedigrees. Tree Genetics & Genomes 6:953-962

Marques C, Brondani R, Grattapaglia D, Sederoff R (2002) Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. Theoretical and Applied Genetics 105:474-478

Martin DM, Aubourg S, Schouwey MB, Daviet L, Schalk M, Toub O, Lund ST, Bohlmann J (2010) Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. BMC Plant Biology 10:226

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evolution 1:vev003-vev003

Martin GB, Bogdanove AJ, Sessa G (2003) Understanding the functions of plant disease resistance proteins. Annual Review of Plant Biology 54:23-61

Martin JT, Juniper BE (1970) The Cuticles of Plants. Edward Arnold, London

McDowell JM, Woffenden BJ (2003) Plant disease resistance genes: recent insights and potential applications. Trends in Biotechnology 21:178-183

McKinnon GE, Jordan GJ, Vaillancourt RE, Steane DA, Potts BM (2004) Glacial refugia and reticulate evolution: the case of the Tasmanian eucalypts. Philosophical Transactions of the Royal Society of London B: Biological Sciences 359:275-284

McTaggart AR, Shuey LS, Granados GM, Plessis Ed, Fraser S, Barnes I, Naidoo S, Wingfield MJ, Roux J (2018) Evidence that *Austropuccinia psidii* may complete its sexual life cycle on Myrtaceae. Plant Pathology 67:729-734

Mendivil-Ramos O, Ferrier DEK (2012) Mechanisms of gene duplication and translocation and progress towards understanding their relative contributions to animal genome evolution. International Journal of Evolutionary Biology 2012:10

Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, Kalinowski J, Linke B, Rupp O, Giegerich R (2003) GenDB—an open source genome annotation system for prokaryote genomes. Nucleic Acids Research 31:2187-2195

Minchinton E, Smith D, Hamley K, Donald C (2014) Myrtle rust in Australia. Acta Horticulturae 1055:89-90

Minh BQ, Nguyen MAT, von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. Molecular Biology and Evolution 30:1188-1195

Moon DH, Salvatierra GR, Caldas DG, de Carvalho MCG, Carneiro RT, Franceschini LM, Oda S, Labate CA (2007) Comparison of the expression profiles of susceptible and resistant *Eucalyptus grandis* exposed to *Puccinia psidii* Winter using SAGE. Functional Plant Biology 34:1010-1018

Morgante M, De Paoli E, Radovic S (2007) Transposable elements and the plant pan-genomes. Current Opinion in Plant Biology 10:149-155

Morin L, Aveyard R, Lidbetter JR, Wilson PG (2012) Investigating the host-range of the rust fungus *Puccinia psidii* sensu lato across tribes of the family Myrtaceae present in Australia. PLoS ONE 7:e35434

Myburg AA, Griffin AR, Sederoff RR, Whetten RW (2003) Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach. Theoretical and Applied Genetics 107:1028-1042

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, Goodstein DM, Dubchak I, Poliakov A, Mizrachi E, Kullan ARK, Hussey SG, Pinard D, Merwe Kvd, Singh P, Jaarsveld Iv, Silva-Junior OB, Togawa RC, Pappas MR, Faria DA, Sansaloni CP, Petroli CD, Yang X, Ranjan P, Tschaplinski TJ, Ye C-Y, Li T, Sterck L, Vanneste K, Murat F, Soler M, Clemente HS, Saidi N, Cassan-Wang H, Dunand C, Hefer CA, Bornberg-Bauer E, Kersting AR, Vining K, Amarasinghe V, Ranik M, Naithani S, Elser J, Boyd AE, Liston A, Spatafora JW, Dharmwardhana P, Raja R, Sullivan C, Romanel E, Alves-Ferreira M, Külheim C, Foley W, Carocha V, Paiva J, Kudrna D, Brommonschenkel SH, Pasquali G, Byrne M, Rigault P, Tibbits J, Spokevicius A, Jones RC, Steane DA, Vaillancourt RE, Potts BM, Joubert F, Barry K, Pappas GJ, Strauss SH, Jaiswal P, Grima-Pettenati J, Salse J, Peer YVd, Rokhsar DS, Schmut J (2014) The genome of *Eucalyptus grandis*. Nature 510:356-362

Myerscough PJ (1998) Ecology of Myrtaceae with special reference to the Sydney region. Cunninghamia 5:787-807

Nagarajan N, Pop M (2013) Sequence assembly demystified. Nature Reviews Genetics 14:157

Neves LG, MC Mamani E, Alfenas AC, Kirst M, Grattapaglia D (2011) A high-density transcript linkage map with 1,845 expressed genes positioned by microarray-based single feature polymorphisms (SFP) in *Eucalyptus*. BMC Genomics 12:1-15

Newman M-A, Von Roepenack E, Daniels M, Dow M (2000) Lipopolysaccharides and plant responses to phytopathogenic bacteria. Molecular Plant Pathology 1:25-31

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32:268-274

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hallman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Kaller M, Luthman J, Lysholm F, Niittyla T, Olson A, Rilakovic N, Ritland C, Rossello JA, Sena J, Svensson T, Talavera-Lopez C, Theiszen G, Tuominen H, Vanneste K, Wu Z-Q, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Garcia Gil R, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Lee Thompson S, Van de Peer Y, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J, Jansson S (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497:579-584

O'Reilly-Wapstra JM, McArthur C, Potts BM (2004) Linking plant genotype, plant defensive chemistry and mammal browsing in a *Eucalyptus* species. Functional Ecology 18:677-684

O'Reilly-Wapstra JM, Freeman JS, Davies NW, Vaillancourt RE, Fitzgerald H, Potts BM (2011) Quantitative trait loci for foliar terpenes in a global eucalypt species. Tree Genetics & Genomes 7:485-498

Oh E, Hansen EM, Sniezko RA (2006) Port-Orford-cedar resistant to *Phytophthora lateralis*. Forest Pathology 36:385-394

Olson MV (1993) The human genome project. Proceedings of the National Academy of Sciences 90:4338-4344

Paap T, Burgess TI, McComb JA, Shearer BL, St J. Hardy GE (2008) *Quambalaria* species, including *Q. coyrecup* sp. nov., implicated in canker and shoot blight diseases causing decline of *Corymbia* species in the southwest of Western Australia. Mycological Research 112:57-69

Padmanabhan MS, Ma S, Burch-Smith TM, Czymmek K, Huijser P, Dinesh-Kumar SP (2013) Novel positive regulatory role for the SPL6 transcription factor in the N TIR-NB-LRR receptor-mediated plant innate immunity. PLoS Pathogens 9:e1003235

Padovan A, Keszei A, Wallis IR, Foley WJ (2012) Mosaic eucalypt trees suggest genetic control at a point that influences several metabolic pathways. Journal of Chemical Ecology 38:914-923

Padovan A, Keszei A, Külheim C, Foley WJ (2014) The evolution of foliar terpene diversity in Myrtaceae. Phytochemistry Reviews 13:695-716

Papantonis A, Cook PR (2013) Transcription factories: genome organization and gene regulation. Chemical Reviews 113:8683-8705

Parkin IAP, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiate DJ (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics 171:765-781

Parra-O. C, Bayly MJ, Drinnan A, Udovicic F, Ladiges P (2009) Phylogeny, major clades and infrageneric classification of *Corymbia* (Myrtaceae), based on nuclear ribosomal DNA and morphology. Australian Systematic Botany 22:384-399

Patharkar OR, Gassmann W, Walker JC (2017) Leaf shedding as an anti-bacterial defense in *Arabidopsis* cauline leaves. PLoS Genetics 13:e1007132

Pegg GS, O'Dwyer C, Carnegie AJ, Burgess TI, Wingfield MJ, Drenth A (2008) *Quambalaria* species associated with plantation and native eucalypts in Australia. Plant Pathology 57:702-714

Pegg GS, Webb R, Carnegie A, Wingfield M, Drenth A (2009) Infection and disease development of *Quambalaria* spp. on *Corymbia* and *Eucalyptus* species. Plant Pathology 58:642-654

Pegg GS, Carnegie AJ, Wingfield MJ, Drenth A (2011) Variable resistance to *Quambalaria pitereka* in spotted gum reveal opportunities for disease screening. Australasian Plant Pathology 40:76-86

Pegg GS, Perry S, Carnegie A, Ireland K, Giblin F (2012) Understanding myrtle rust epidemiology and host specificity to determine disease impact in Australia. Cooperative Research Centre for National Plant Biosecurity, Bruce, Australia

Pegg GS, Brawner JT, Lee DJ (2014a) Screening *Corymbia* populations for resistance to *Puccinia psidii*. Plant Pathology 63:425-436

Pegg GS, Giblin F, McTaggart A, Guymer G, Taylor H, Ireland K, Shivas R, Perry S (2014b) *Puccinia psidii* in Queensland, Australia: disease symptoms, distribution and impact. Plant Pathology 63:1005-1021

Pegg GS, Taylor T, Entwistle P, Guymer G, Giblin F, Carnegie A (2017) Impact of *Austropuccinia psidii* (myrtle rust) on Myrtaceae-rich wet sclerophyll forests in south east Queensland. PloS One 12:e0188058

Peñuelas J, Llusià J, Asensio D, Munné-Bosch S (2005) Linking isoprene with plant thermotolerance, antioxidants and monoterpene emissions. Plant, Cell & Environment 28:278-286

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. Annual Review of Ecology, Evolution, and Systematics 37:187-214

Petroli CD, Sansaloni CP, Carling J, Steane DA, Vaillancourt RE, Myburg AA, da Silva OB, Jr., Pappas GJ, Jr., Kilian A, Grattapaglia D (2012) Genomic characterization of DArT markers based on high-density linkage analysis and physical mapping to the *Eucalyptus* genome. PLoS ONE 7:e44684

Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. Molecular Breeding 7:275-291

Pichersky E, Gershenzon J (2002) The formation and function of plant volatiles: perfumes for pollinator attraction and defense. Current Opinion in Plant Biology 5:237-243

Pinkard E, Kriticos D, Wardlaw T, Carnegie A, Leriche A (2010) Estimating the spatio-temporal risk of disease epidemics using a bioclimatic niche model. Ecological Modelling 221:2828-2838

Piza SdT, Ribeiro I (1988) Influência da luz e da temperatura na germinação de uredosporos de *Puccinia psidii*. Bragantia 47:75-78

Poethig RS (2009) Small RNAs and developmental timing in plants. Current Opinion in Genetics & Development 19:374-378

Poke FS, Vaillancourt RE, Potts BM, Reid JB (2005) Genomic research in *Eucalyptus*. Genetica 125:79-101

Potts BM, Barbour RC, Hingston AB, Vaillancourt RE (2003) Genetic pollution of native eucalypt gene pools - identifying the risks. Australian Journal of Botany 51:1-25

Potts BM, Dungey HS (2004) Interspecific hybridization of *Eucalyptus*: key issues for breeders and geneticists. New Forests 27:115-138

Potts BM, Vaillancourt RE, Jordan G, Dutkowski G, Costa e Silva J, McKinnon G, Steane D, Volker P, Lopez G, Apiolaza L, Li Y, Marques C, Borralho N (2004) Exploration of the *Eucalyptus globulus* gene pool. In: Borralho N, Pereira J, Marques C, Coutinho J, Madeira M, Tomé M (eds) *Eucalyptus* in a changing world. Proceedings of an IUFRO conference, Aveiro, Portugal. RAIZ, Instituto Investigação de Floresta e Papel, pp 46-61

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rae AM, Street NR, Robinson KM, Harris N, Taylor G (2009) Five QTL hotspots for yield in short rotation coppice bioenergy poplar: the poplar biomass loci. BMC Plant Biology 9:23

Remington DL, Ungerer M, Purugganan M (2001) Map-based cloning of quantitative trait loci: progress and prospects. Genetics Research 78:213-218

Ren Y, Zhao H, Kou Q, Jiang J, Guo S, Zhang H, Hou W, Zou X, Sun H, Gong G (2012) A high resolution genetic map anchoring scaffolds of the sequenced watermelon genome. PLoS One 7:e29453

Reuter Jason A, Spacek DV, Snyder Michael P (2015) High-throughput sequencing technologies. Molecular Cell 58:586-597

Rizzo DM, Garbelotto M (2003) Sudden oak death: endangering California and Oregon forest ecosystems. Frontiers in Ecology and the Environment 1:197-204

Rockwood D, Rudie A, Ralph S, Zhu J, Winandy J (2008) Energy product options for *Eucalyptus s*pecies grown as short rotation woody crops. International Journal of Molecular Sciences 9:1361

Rosado CCG, da Silva Guimarães LM, Faria DA, de Resende MDV, Cruz CD, Grattapaglia D, Alfenas AC (2016) QTL mapping for resistance to Ceratocystis wilt in *Eucalyptus*. Tree Genetics & Genomes 12:72

Rosado TB, Tomaz RS, Ribeiro Junior MF, Rosado AM, Guimarães LMdS, Araújo EFd, Alfenas AC, Cruz CD (2010) Detection of QTL associated with rust resistance using IBD-based methodologies in exogamic *Eucalyptus* spp. populations. Crop Breeding and Applied Biotechnology 10:321-328

Roth C, Rastogi S, Arvestad L, Dittmar K, Light S, Ekman D, Liberles DA (2007) Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 308B:58-73

Roux J, Mthalane Z, De Beer Z, Eisenberg B, Wingfield M (2006) Quambalaria leaf and shoot blight on *Eucalyptus nitens* in South Africa. Australasian Plant Pathology 35:427-433

Roux J, Greyling I, Coutinho TA, Verleur M, Wingfield MJ (2013) The myrtle rust pathogen, *Puccinia psidii*, discovered in Africa. IMA Fungus 4:155

Saghai Maroof MA, Yang GP, Biyashev RM, Maughan PJ, Zhang Q (1996) Analysis of the barley and rice genomes by comparative RFLP linkage mapping. Theoretical and Applied Genetics 92:541-551

Sakaguchi S, Sugino T, Tsumura Y, Ito M, Crisp MD, Bowman DMJS, Nagano AJ, Honjo MN, Yasugi M, Kudoh H, Matsuki Y, Suyama Y, Isagi Y (2015) High-throughput linkage mapping of Australian white cypress pine (*Callitris glaucophylla*) and map transferability to related species. Tree Genetics & Genomes 11:1-12

Sakamoto T, Danzmann RG, Gharbi K, Howard P, Ozaki A, Khoo SK, Woram RA, Okamoto N, Ferguson MM, Holm L-E, Guyomard R, Hoyheim B (2000) A microsatellite linkage

map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. Genetics 155:1331-1345

Salgado H, Moreno-Hagelsieb G, Smith TF, Collado-Vides J (2000) Operons in *Escherichia coli*: genomic analyses and predictions. Proceedings of the National Academy of Sciences 97:6652-6657

Salse J (2012) In silico archeogenomics unveils modern plant genome organisation, regulation and evolution. Current Opinion in Plant Biology 15:122-130

Sambrook J, Williams J, Sharp PA, Grodzicker T (1975) Physical mapping of temperature-sensitive mutations of adenoviruses. Journal of Molecular Biology 97:369-390

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 74:5463-5467

Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A (2011) Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. BMC Proceedings 5:1-2

Santos MR, da Silva Guimarães LM, de Resende MDV, Rosse LN, Zamprogno KC, Alfenas AC (2014) Eucalypts rust (*Puccinia psidii*) resistance in *Eucalyptus pellita*. Crop Breeding and Applied Biotechnology 14:244-250

Sasaki K, Saito T, Lämsä M, Oksman-Caldentey K-M, Suzuki M, Ohyama K, Muranaka T, Ohara K, Yazaki K (2007) Plants utilize isoprene emission as a thermotolerance mechanism. Plant and Cell Physiology 48:1254-1262

Schiffer PH, Gravemeyer J, Rauscher M, Wiehe T (2016) Ultra large gene families: a matter of adaptation or genomic parasites? Life 6:32

Schnee C, Köllner TG, Held M, Turlings TCJ, Gershenzon J, Degenhardt J (2006) The products of a single maize sesquiterpene synthase form a volatile defense signal that attracts natural enemies of maize herbivores. Proceedings of the National Academy of Sciences of the United States of America 103:1129-1134

Schneeberger K, Ossowski S, Ott F, Klein JD, Wang X, Lanz C, Smith LM, Cao J, Fitz J, Warthmann N (2011) Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. Proceedings of the National Academy of Sciences 108:10249-10254

Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C (2011) Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nature Genetics 43:333

Schwab W, Fuchs C, Huang F-C (2013) Transformation of terpenes into fine chemicals. European Journal of Lipid Science and Technology 115:3-8

Semagn K, Bjørnstad Å, Ndjiondjop M (2006) An overview of molecular marker methods for plants. African Journal of Biotechnology 5:2540–2568

Sewell MM, Neale DB (2000) Mapping quantitative traits in forest trees. In: Jain SM, Minocha SC (eds) Molecular Biology of Woody Plants: Volume 1. Springer Netherlands, Dordrecht, pp 407-423. doi:10.1007/978-94-017-2311-4_17

Sharma M, Pandey GK (2015) Expansion and function of repeat domain proteins during stress and development in plants. Frontiers in Plant Science 6:1218

Shepherd M, Chaparro JX, Teasdale R (1999) Genetic mapping of monoterpene composition in an interspecific eucalypt hybrid. Theoretical and Applied Genetics 99:1207-1215

Shepherd M, Baker N, Kasem S, Lee DJ, Henry RJ (2006) Comparative mapping of *Eucalyptus* and *Corymbia*. In: Breeding for success: diversity in action: Proceedings of the 13th Australasian Plant Breeding Conference, Dunedin, New Zealand.

Shepherd M, Bartle J, Lee DJ, Brawner J, Bush D, Turnbull P, Macdonel P, Brown TR, Simmons B, Henry R (2011) Eucalypts as a biofuel feedstock. Biofuels 2:639-657

Shepherd M, Baten A, Junior OBdS, Lee DJ, Butler JB, Freeman J, Vaillancourt R, Potts B, Grattapaglia D, King G (2015) Towards a *Corymbia* reference genome: comparative efficiencies of Illumina, PacBio and hybrid *de novo* assemblies of a complex heterozygous genome. In: Vettori C, Vendramin GG, Paffetti D, Travaglini D (eds) Proceedings of the IUFRO Tree Biotechnology 2015 Conference: "Forests: the importance to the planet and society", Florence, Italy.

Shepherd M, Barry KW, Baten A, Butler JB, Freeman JS, Furtado A, Grattapaglia D, Healey A, Henry RJ, King GJ, Lee D, Potts BM, Schmutz J, da Silva Junior OB, Simmons B, Vaillancourt R (2016) Spotting the difference: comparing the genome of *Corymbia* with its larger cousin *Eucalyptus grandis*. In: Plant & Animal Genome Conference XXIV, San Diego, CA.

Shete S, Tiwari H, Elston RC (2000) On estimating the heterozygosity and polymorphism information content value. Theoretical Population Biology 57:265-271

Silva-Junior OB, Grattapaglia D (2015) Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. New Phytologist 208:830-845

Simonin KA, Roddy AB (2018) Genome downsizing, physiological novelty, and the global dominance of flowering plants. PLoS Biology 16:e2003706

Simpson JA (2000) *Quambalaria*, a new genus of eucalypt pathogens. Australasian Mycologist 19:57-62

Singsaas EL, Lerdau M, Winter K, Sharkey TD (1997) Isoprene increases thermotolerance of isoprene-emitting species. Plant Physiology 115:1413-1420

Sinnott EW (1916) Comparative rapidity of evolution in various plant types. The American Naturalist 50:466-478

Slee A, Brooker M, Duffy S, West J (2006) EUCLID eucalypts of Australia. 3rd edn. Centre for Plant Biodiversity Research - CSIRO Publishing Canberra

Smith A, Potts B, Ratkowsky D, Pinkard E, Mohammed C (2017) Association of *Eucalyptus globulus* leaf anatomy with susceptibility to *Teratosphaeria* leaf disease. Forest Pathology 48:e12395

Soler M, Camargo ELO, Carocha V, Cassan-Wang H, San Clemente H, Savelli B, Hefer CA, Paiva JAP, Myburg AA, Grima-Pettenati J (2015) The *Eucalyptus grandis* R2R3-MYB transcription factor family: evidence for woody growth-related evolution and function. New Phytologist 206:1364-1377

Soltis DE, Visger CJ, Soltis PS (2014) The polyploidy revolution then…and now: Stebbins revisited. American Journal of Botany 101:1057-1078

Stanke M, Tzvetkova A, Morgenstern B (2006) AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. Genome Biology 7:S11

Stebbins C (1950) Variation and evolution in plants. Columbia University Press, New York and London

Sterck L, Rombauts S, Jansson S, Sterky F, Rouzé P, Van de Peer Y (2005) EST data suggest that poplar is an ancient polyploid. New Phytologist 167:165-170

Stone C, Simpson JA, Eldridge RH (1998) Insect and fungal damage to young eucalypt trial plantings in northern New South Wales. Australian Forestry 61:7-20

Strauss SH, Myburg AA (2015) Plant scientists celebrate new woody plant genome. New Phytologist 206:1185-1187

Sturtevant A (1913) The linear arrangement of six X-linked factors in *Drosophila*, as shown by their mode of association. Journal of Experimental Zoology 14:43-59

Sun Z, Wang Z, Tu J, Zhang J, Yu F, McVetty PBE, Li G (2007) An ultradense genetic recombination map for *Brassica napus*, consisting of 13551 SRAP markers. Theoretical and Applied Genetics 114:1305-1317

Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004) Close split of sorghum and maize genome progenitors. Genome Research 14:1916-1923

Szöllősi GJ, Daubin V (2012) Modeling gene family evolution and reconciling phylogenetic discord. Evolutionary Genomics: Statistical and Computational Methods 2:29-51

Takos AM, Rook F (2012) Why biosynthetic genes for chemical defense compounds cluster. Trends in Plant Science 17:383-388

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008) Synteny and collinearity in plant genomes. Science 320:486-488

Tang Y-C, Amon A (2013) Gene copy-number alterations: a cost-benefit analysis. Cell 152:394-405

Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Research 12:4127-4138

Taylor AL, Trotter CD (1967) Revised linkage map of *Escherichia coli*. Bacteriological Reviews 31:332

Taylor JS, Van de Peer Y, Meyer A (2001) Genome duplication, divergent resolution and speciation. Trends in Genetics 17:299-301

Tennessen JA, Govindarajulu R, Ashman T-L, Liston A (2014) Evolutionary origins and dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps. Genome Biology and Evolution 6:3295-3313

Thavamanikumar S, McManus LJ, Ades PK, Bossinger G, Stackpole DJ, Kerr R, Hadjigol S, Freeman JS, Vaillancourt RE, Zhu P, Tibbits JFG (2014) Association mapping for wood quality and growth traits in *Eucalyptus globulus* ssp. *globulus* Labill identifies nine stable marker-trait associations for seven traits. Tree Genetics & Genomes 10:1661-1678

Thomma BP, Nürnberger T, Joosten MH (2011) Of PAMPs and effectors: the blurred PTI-ETI dichotomy. Plant Cell 23:4-15

Thornhill AH, Ho SYW, Külheim C, Crisp MD (2015) Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. Molecular Phylogenetics and Evolution 93:29-43

Thumma BR, Southerton SG, Bell JC, Owen JV, Henery ML, Moran GF (2010) Quantitative trait locus (QTL) analysis of wood quality traits in *Eucalyptus nitens*. Tree Genetics & Genomes 6:305-317

Thumma BR, Pegg GS, Warburton P, Brawner J, Macdonell P, Yang X, Southerton S (2013) Molecular tagging of rust resistance genes in eucalypts. CSIRO Plant Industry, Canberra

Thumma BR, Thavamanikumar S, Southerton S (2017) Discovery and application of DNA markers for resistance to *Teratosphaeria* in *E. globulus*. Forest & Wood Products Australia, Melbourne

Tobias PA, Park RF, Külheim C, Guest DI (2015) Wild-sourced *Chamelaucium uncinatum* have no resistance to *Puccinia psidii* (myrtle rust). Australasian Plant Disease Notes 10:15

Tobias PA, Guest DI, Külheim C, Hsieh JF, Park RF (2016) A curious case of resistance to a new encounter pathogen: myrtle rust in Australia. Molecular Plant Pathology 17:783-788

Tobias PA, Guest D, Külheim C, Park RF (2017) *De novo* transcriptome study identifies candidate genes involved in resistance to *Austropuccinia psidii* (myrtle rust) in *Syzygium luehmannnii* (Riberry). Phytopathology 108:627-640

Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635-641

Treangen TJ, Salzberg SL (2012) Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nature Reviews Genetics 13:36

Tsuda K, Somssich IE (2015) Transcriptional networks in plant immunity. New Phytologist 206:932-947

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP,

Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen G-L, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé J-C, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai C-J, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596-1604

Uchida J, Zhong S, Killgore E (2006) First report of a rust disease on ohia caused by *Puccinia psidii* in Hawaii. Plant Disease 90:524-524

Van Der Biezen EA, Jones JD (1998) Plant disease-resistance proteins and the gene-for-gene concept. Trends in Biochemical Sciences 23:454-456

Van der Colff D, Dreyer LL, Valentine A, Roets F (2017) Differences in physiological responses to infection by *Ceratocystis tsitsikammensis*, a native ophiostomatoid pathogen, between a native forest and an exotic forestry tree in South Africa. Fungal Ecology 27:107-115

Van der Hoorn RAL, Kamoun S (2008) From guard to decoy: a new model for perception of plant pathogen effectors. The Plant Cell 20:2009-2017

Van Ooijen J (2006) JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. *Kyazma BV*, Wageningen, The Netherlands

Van Ooijen J (2009) MapQTL 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma BV: Wageningen, Netherlands

Veeckman E, Ruttink T, Vandepoele K (2016) Are we there yet? Reliably estimating the completeness of plant genome sequences. The Plant Cell 28:1759-1768

Veitia RA (2004) Gene dosage balance in cellular pathways: implications for dominance and gene duplicability. Genetics 168:569

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA (2001) The sequence of the human genome. Science 291:1304-1351

Vernin GA, Parkanyi C, Cozzolino F, Fellous R (2004) GC/MS analysis of the volatile constituents of *Corymbia citriodora* Hook. from Réunion Island. Journal of Essential Oil Research 16:560-565

Vinckenbosch N, Dupanloup I, Kaessmann H (2006) Evolutionary fate of retroposed gene copies in the human genome. Proceedings of the National Academy of Sciences of the United States of America 103:3220-3225

Wallace HM, Trueman SJ (1995) Dispersal of *Eucalyptus torelliana* seeds by the resin-collecting stingless bee, *Trigona carbonaria*. Oecologia 104:12-16

Wallace HM, Howell MG, Lee DJ (2008) Standard yet unusual mechanisms of long-distance dispersal: seed dispersal of *Corymbia torelliana* by bees. Diversity and Distributions 14:87-94

Wang S, Chen J, Zhang W, Hu Y, Chang L, Fang L, Wang Q, Lv F, Wu H, Si Z, Chen S, Cai C, Zhu X, Zhou B, Guo W, Zhang T (2015) Sequence-based ultra-dense genetic and physical maps reveal structural variations of allopolyploid cotton genomes. Genome Biology 16:108

Wang Y, Wang X, Paterson AH (2012) Genome and gene duplications and gene expression divergence: a view from plants. Annals of the New York Academy of Sciences 1256:1-14

Wang Z, Taramino G, Yang D, Liu G, Tingey S, Miao G, Wang G (2001) Rice ESTs with disease-resistance gene-or defense-response gene-like sequences mapped to regions containing major resistance genes or QTLs. Molecular Genetics and Genomics 265:302-310

Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B (2016) gplots: various R programming tools for plotting data. https://CRAN.R-project.org/package=gplots

Wenke K, Kai M, Piechulla B (2010) Belowground volatiles facilitate interactions between plant roots and soil organisms. Planta 231:499-506

Wenzel G, Foroughi-Wehr B (1990) Progeny tests of barley, wheat, and potato regenerated from cell cultures after in vitro selection for disease resistance. Theoretical and Applied Genetics 80:359-365

Westaway JO (2016) The pathogen Myrtle Rust (*'Puccinia psidii'*) in the Northern Territory: first detection, new host and potential impacts. Northern Territory Naturalist 27:13

Wikström N, Savolainen V, Chase MW (2001) Evolution of the angiosperms: calibrating the family tree. Proceedings of the Royal Society of London Series B: Biological Sciences 268:2211-2220

Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM (2009) Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. Plant Physiology 149:981-993

Wilkinson J (1944) The cytology of *Salix* in relation to its taxonomy. Annals of Botany 8:269-284

Williams DJ, Dancik BP, Pharis RP (1987) Early progeny testing and evaluation of controlled crosses of black spruce. Canadian Journal of Forest Research 17:1442-1450

Williams EJ, Bowles DJ (2004) Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*. Genome Research 14:1060-1067

Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18:6531-6535

Wilson C, Zhan H, Swint-Kruse L, Matthews K (2007) The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. Cellular and Molecular Life Sciences 64:3-16

Wingfield MJ, Crous PW, Swart WJ (1993) *Sporothrix eucalypti* (sp. nov.), a shoot and leaf pathogen of *Eucalyptus* in South Africa. Mycopathologia 123:159-164

Wu G, Park MY, Conway SR, Wang J-W, Weigel D, Poethig RS (2009) The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. Cell 138:750-759

Wu R (1998) Genetic mapping of QTLs affecting tree growth and architecture in *Populus*: implication for ideotype breeding. Theoretical and Applied Genetics 96:447-457

Xavier AA, Alfenas AC, Matsuoka K, Hodges CS (2001) Infection of resistant and susceptible *Eucalyptus grandis* genotypes by urediniospores of *Puccinia psidii*. Australasian Plant Pathology 30:277-281

Xiao S, Wang W, Yang X (2008) Evolution of resistance genes in plants. In: Heine H (ed) Innate Immunity of Plants, Animals, and Humans, vol 21. Nucleic Acids and Molecular Biology. Springer Berlin Heidelberg, pp 1-25

Xiong W, Wu P, Jia Y, Wei X, Xu L, Yang Y, Qiu D, Chen Y, Li M, Jiang H, Wu G (2016) Genome-wide analysis of the terpene synthase gene family in physic nut (*Jatropha curcas* L.) and functional identification of six terpene synthases. Tree Genetics & Genomes 12:97

Yang S, Zhang X, Yue J-X, Tian D, Chen J-Q (2008) Recent duplications dominate NBS-encoding gene expansion in two woody species. Molecular Genetics and Genomics 280:187-198

Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration–selection balance. Evolution 65:1897-1911

Yuskianti V, Glen M, Puspitasari D, Francis A, Rimbawanto A, Gafur A, Indrayadi H, Mohammed C (2014) Species-specific PCR for rapid identification of *Ganoderma philippii* and *Ganoderma mastoporum* from *Acacia mangium* and *Eucalyptus pellita* plantations in Indonesia. Forest Pathology 44:477-485

Zauza EAV, Alfenas AC, Old K, Couto MMF, Graça RN, Maffia LA (2010) Myrtaceae species resistance to rust caused by *Puccinia psidii*. Australasian Plant Pathology 39:406-411

Zeng Z-B, Liu J, Stam LF, Kao C-H, Mercer JM, Laurie CC (2000) Genetic architecture of a morphological shape difference between two *Drosophila* species. Genetics 154:299-310

Zhai J, Jeong D-H, De Paoli E, Park S, Rosen BD, Li Y, González AJ, Yan Z, Kitto SL, Grusak MA (2011) MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. Genes & Development 25:2540-2553

Zhang B, Pan X, Cobb GP, Anderson TA (2006) Plant microRNA: a small regulatory molecule with big impact. Developmental Biology 289:3-16

Zhang S-D, Ling L-Z, Yi T-S (2015) Evolution and divergence of SBP-box genes in land plants. BMC Genomics 16:787

Zhao J-P, Jiang X-L, Zhang B-Y, Su X-H (2012) Involvement of microRNA-mediated gene expression regulation in the pathological development of stem canker disease in *Populus trichocarpa*. PLoS ONE 7:e44968

Zheng D, Gerstein MB (2007) The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they? Trends in Genetics 23:219-224

Zhong S, Yang B, Puri K (2011) Characterization of *Puccinia psidii* isolates in Hawaii using microsatellite DNA markers. Journal of General Plant Pathology 77:178-181

Zhou X, De Beer ZW, Xie Y, Pegg GS, Wingfield MJ (2007) DNA-based identification of *Quambalaria pitereka* causing severe leaf blight of *Corymbia citriodora* in China. Fungal Diversity 25:245-254

Zhuang J-Y, Wei S-X (2011) Additional materials for the rust flora of Hainan Province, China. Mycosystema 30:853-860

Żmieńko A, Samelak A, Kozłowski P, Figlerowicz M (2014) Copy number polymorphism in plant genomes. Theoretical and Applied Genetics 127:1-18

Zobel B (1993) Clonal forestry in the eucalypts. In: Clonal Forestry II. Springer, pp 139-148

Zou F, Chai HS, Younkin CS, Allen M, Crook J, Pankratz VS, Carrasquillo MM, Rowley CN, Nair AA, Middha S (2012) Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. PLoS Genetics 8:e1002707